(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2001 (26.04.2001)

PCT

(10) International Publication Number
WO 01/29211 A2

(51) International Patent Classification[7]: C12N 15/10, C12Q 1/68, C12P 19/34

(21) International Application Number: PCT/US00/29046

(22) International Filing Date: 19 October 2000 (19.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/160,420	19 October 1999 (19.10.1999)	US
09/514,660	29 February 2000 (29.02.2000)	US
60/218,883	18 July 2000 (18.07.2000)	US
60/219,087	18 July 2000 (18.07.2000)	US
60/219,090	18 July 2000 (18.07.2000)	US
09/618,696	18 July 2000 (18.07.2000)	US
09/618,935	18 July 2000 (18.07.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

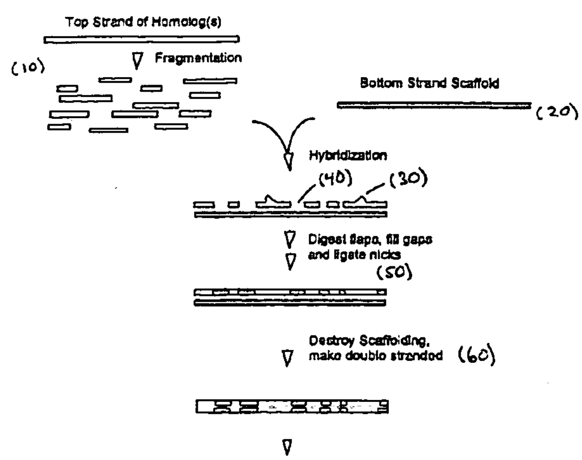
US 09/514,660 (CIP)
Filed on 29 February 2000 (29.02.2000)
US 09/618,696 (CIP)
Filed on 18 July 2000 (18.07.2000)
US 09/618,935 (CIP)
Filed on 18 July 2000 (18.07.2000)

(71) Applicant (*for all designated States except US*): **ENCHIRA BIOTECHNOLOGY CORPORATION** [US/US]; 4200 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **COCO, Wayne, M.** [US/US]; 43 Cornerbrook Place, The Woodlands, TX 77381 (US). **ARENSDORF, Joseph, J.** [US/US]; 27225 Orth Lane, Oak Ridge North, TX 77385 (US). **CRIST, Michael, J.** [US/US]; 1638 Woodcrest Drive, Houston, TX 77018 (US). **DARZINS, Aldis** [US/US]; 23 Firethorn Place, The Woodlands, TX 77381 (US). **ENCELL, Lance, P.** [US/US]; 58 Brooksedge Circle, The Woodlands, TX 77382 (US). **FOLSOM, Brian, R.** [US/US]; 94 West Elm Crescent, The Woodlands, TX 77382 (US). **HEKTOR, Hans, J.** [NL/US]; 2301 S. Millbend Drive #112, The Woodlands, TX 77380 (US). **NALTY, Mark, S.** [US/US]; 1600 River Pointe #1172, Conroe, TX 77304 (US). **LEVINSON, William, E.** [US/US]; 15 Fallshire Drive,

[Continued on next page]

(54) Title: METHOD FOR DIRECTED EVOLUTION BY RANDOM CHIMERAGENESIS ON TRANSIENT TEMPLATES



(57) Abstract: A method is provided wherein at least one template is used to align oligonucleotides of interest or fragments thereof to generate chimeric polynucleotides. The oligonucleotides are allowed to hybridize to the template or templates. If necessary, unhybridized termini of hybridized oligonucleotides are removed and gaps between hybridized oligonucleotides are filled. Adjacently hybridized oligonucleotides are ligated to generate chimeric polynucleotides hybridized to the template or templates. The chimeric polynucleotides can then be subjected to selection and/or screening for desired characteristics. Prior to or concomitant with selection/screening, the template is optionally removed.

WO 01/29211 A2

The Woodlands, TX 77381 (US). **PALMER, Stephen, R.** [GB/US]: 8 Greentwig Place, The Woodlands, TX 77381 (US).

(74) Agents: ELMORE, Carolyn, S. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD FOR DIRECTED EVOLUTION BY RANDOM CHIMERAGENESIS ON TRANSIENT TEMPLATES

RELATED APPLICATIONS

This application is a continuation-in-part of United States Application No. 09/618,935 filed July 18, 2000 and a continuation-in-part of United States Application 09/618,696 filed July 18, 2000 and a continuation-in-part of United States Application No. 09/514,660 filed February 29, 2000, which claims the benefit of United States Provisional Application No. 60/160,420, filed October 19, 1999. This application also claims the benefit of United States Provisional Application 60/219,090 filed July 18, 2000 and United States Provisional Application 60/218,883 filed July 18, 2000 and United States Provisional Application 60/219,087 filed July 18, 2000. The teachings of all the cited applications are incorporated by reference herein in their entireties.

BACKGROUND OF THE INVENTION

Genetic improvements occur more frequently when the generation of mutations is coupled with genetic recombination. The effect of genetic recombination on the fixing in a population of multiple beneficial mutations is evident when comparing sexually versus asexually replicating organisms. Asexually replicating organisms exhibit an accumulation of mutations that limit their

evolutionary potential because they have no mechanism to combine beneficial mutations while eliminating detrimental mutations due to the fact that genetic recombination does not occur. This reduction in evolutionary potential in asexually replicating populations is known as Müller's ratchet (Müller, H., *Mut. Res. 1*:2-9, 1964). The process of altering genetic functions through generation of mutants, and/or chimeric genetic recombinants, coupled with selection and/or screening is termed "directed evolution." Recombination between altered or otherwise non-identical polynucleotide targets allows the consolidation of favorable mutations that originally occurred on separate copies of the target, as well as the elimination of detrimental mutations (Harayama, S. *Trends Biotechnol.*, 16:76-82 (1998)).

Methods of directed evolution include the "sexual PCR" method, which involves cleaving a population of target DNA using DNase I, followed by the reassembly/recombination of DNA fragments by a self-primed Polymerase Chain Reaction (hereinafter "PCR") (W.P.C. Stemmer, (1994) *Nature, 370*:389-391; United States Patent Nos. 5,605,793 and 5,811,238).

Alternative methods for directed evolution involve modifications of the sexual PCR method such as "Random Priming Recombination" (hereinafter "RPR"; Shao, Z. *et al.*, *Nucleic Acids Res. 26*:681-3, 1998), and other methods, such as the "Staggered Extension Process" (hereinafter "StEP"; Zhao, H. *et al.*, *Nat. Biotechnol. 16*:258-61, 1998). An essential step for both of these methods is a PCR step. RPR utilizes elongation of randomly annealed primers to synthesize fragments from a given template. These fragments are then used for the reassembly step in the sexual PCR method. StEP involves the binding of a primer to the terminus of a population of a mutant or wild-type template. A brief round of polymerization extends the primers, however, extension is prematurely halted, resulting in a DNA fragment shorter than the original template. The short fragment is denatured from its original template and allowed to reanneal, potentially to a template with a different mutation. Subsequent brief rounds of polymerization and denaturation further extend these fragments, eventually producing newly synthesized strands as long as the template strand.

The result of these methods of directed evolution are "chimeric" polynucleotides, so called because they include sequences from more than one parent gene. Sexual PCR, RPR and StEP each result in a library of recombinant DNA chimeras of the starting parental genes, however, due to limitations inherent in the essential PCR step in each of these methods, the library of chimeric products generally represent a limited sampling of all potential chimeric products.

Specifically, in sexual PCR, priming among fragments of the same gene is favored, thus decreasing the number of crossovers and increasing the proportion of parents recovered in recombinant libraries (Kikuchi, M., *et al.*, Gene 243:133-137, 2000). This is particularly the case when few homologous genes are available, or desired, for a given reaction. Merely generating or screening larger libraries cannot compensate for such lack of diversity, since the diversity inherent in the family shuffling of even two average-sized genes which are ninety percent identical is more than a dozen orders of magnitude larger than can be captured in any recombinant library. Thus, each library clone with few or no crossovers limits the exploration of the sequence diversity contained in more highly mosaic clones.

Moreover, these methods can suffer from "blind spots" in the gene or polypeptide of interest, where exchanges between parental DNA from two or more sources is rare or nonexistent due to the manner in which the DNA is fragmented or because regions of homology of a certain size are generally required to allow homologous recombination between the parental DNA.

SUMMARY OF THE INVENTION

The method of the present invention facilitates the generation of chimeric polynucleotides. "Chimeric polynucleotides," as used herein, contain nucleotide sequences from multiple sources, typically related sequences or otherwise similar polynucleotides, referred to herein as "parent polynucleotides". Any number of parent polynucleotides can be utilized. In particular embodiments two, three and four parents are present. The partial sequences are usually presented in the form of an "oligonucleotide population". The method facilitates the recombination of regions of parent polynucleotides derived or isolated from any suitable source of

nucleic acid, including known sequences, unknown or uncharacterized sequences, *e.g.*, nucleic acid molecules isolated from environmental or other sources, and regions of randomly generated sequence, into a chimeric polynucleotide. In addition, the method allows for deletion and/or insertion of sequences from reference parent polynucleotides compared to the resultant chimeric polynucleotides.

The present method, termed "Random Chimeragenesis on Transient Templates" (hereinafter the "RACHITT™ method") results in low levels of siblings (clones produced by replication rather than separate recombination events), inactive proteins or unshuffled parental clones. The method also yields an unprecedented predominance of highly chimeragenic (mosaic) clones and provides the unique ability to effect frequent recombination between close and even adjacent alleles. These observed improvements are in the key mechanisms that determine the diversity and evolutionary potential of gene shuffled libraries. These elements combined indicate an ability to explore more diverse permutations of multiple mutations with the RACHITT™ method than has been possible using other methods. The ability to capture untapped regions of sequence space better allows the evolution of polynucleotides encoding proteins having desired characteristics.

While not wishing to be bound by theory, the advantages of the RACHITT™ method can be attributed, at least in part, to the aspects of the method as described below. The RACHITT™ method utilizes a single round of hybridization, as opposed to reiterative thermocycling, and thus lends itself to greater specificity in similar or more divergent pairings. When necessary, it also incorporates "flap" trimming to result in perfectly duplexed ligatable/extendable ends from otherwise unproductive fragments. Trimming also allows incorporation of fragments much smaller that those generated by DNase I. In addition, many DNA ligases allow ligation of mismatched termini. Where flaps are not perfectly trimmed, such low-fidelity ligation may contribute to incorporation of fragments with adjacent mismatches. Ligation of adjacent fragments from the same parent also allows linking of larger stretches from any given parent than were present after DNase I cleavage. Finally, the RACHITT™ method typically exploits a bottom strand template from one parent

and top strand fragments from the same and/or other parents. This prevents parental fragments from reannealing to their own complementary strands.

In one aspect, the present invention is drawn to a method for forming at least one chimeric polynucleotide including the steps of contacting a single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template and treating the hybridized oligonucleotides to form a chimeric polynucleotide. The population of oligonucleotides can be random. The method can further include steps of trimming flaps and filling in gaps.

In another embodiment, the invention is directed to a method for forming a chimeric polynucleotide including contacting a single-stranded polynucleotide template with a population of oligonucleotides under conditions such that at least two oligonucleotides hybridize to a given template, and wherein the population of oligonucleotides comprises oligonucleotides such that two or more regions of the template are complementary to two or more oligonucleotides of the population; and ligating the hybridized oligonucleotides such that one chimeric polynucleotide, is generated.

In another embodiment, the invention is directed to a method for forming at least one chimeric polynucleotide including contacting a single-stranded polynucleotide template with a population of oligonucleotides, wherein at least two of the oligonucleotides hybridize to the same template, and containing at least one flap; removing flaps; and ligating the hybridized oligonucleotides such that one chimeric polynucleotide, is generated.

In another embodiment of the present invention is drawn to a method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric polynucleotide is altered in comparison to at least one reference polynucleotide. This embodiment includes the steps of generating a population of oligonucleotides, comprising randomly fragmenting at least one polynucleotide; contacting at least one single-stranded polynucleotide template with the population of oligonucleotides under conditions wherein at least two oligonucleotides hybridize to the template; optionally removing flaps filling in gaps between hybridized

oligonucleotides; ligating hybridized oligonucleotides to form at least one chimeric polynucleotide, hybridized to a template; selectively amplifying the chimeric polynucleotides and selecting or screening at least one chimeric polynucleotide having said specified characteristic altered in comparison to the reference polynucleotide.

In another embodiment, the invention is directed to a method for forming a chimeric polynucleotide including contacting a single-stranded template with a random population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template, filling in gaps between the hybridized oligonucleotides, and ligating the hybridized oligonucleotides such that a chimeric polynucleotide is formed. The populations of oligonucleotides can be random. The method can further include a step of trimming flaps. Both 3' and 5' flaps can be trimmed. The template can include a coding region. A template-length chimeric polynucleotide can be formed. The population of oligonucleotides can be produced by fragmenting a single-stranded nucleic acid. Alternatively, the population of oligonucleotides can be produced synthetically. A plurality of chimeric polynucleotides can be formed. The number of chimeric polynucleotides formed and the number of single-stranded templates can be in a ratio of about 1.

In another embodiment, the invention is directed to a non-reiterative process for forming a template-length chimeric polynucleotide including contacting a single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide. The method can further include a step of filling-in gaps. The method can further include a step of trimming flaps.

In another embodiment, the invention is directed to a method for forming a chimeric polynucleotide including the steps of contacting a single-stranded template with a population of oligonucleotides produced by fragmenting a single-stranded nucleic acid or by chemical synthesis under conditions such that at least two of the oligonucleotides hybridize to the template; and ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide. The

population of oligonucleotides can be random. The method can further include a step of filling-in gaps. The method can further include a step of trimming flaps.

In still another embodiment, the invention is directed to a method for forming a plurality of chimeric polynucleotides on single-stranded polynucleotide templates, wherein the number of chimeric polynucleotides formed and the number of single-stranded templates is in a ratio of about 1 including the steps of contacting a single-stranded template with a population of oligonucleotides produced by fragmenting a single-stranded nucleic acid or by chemical synthesis, under conditions such that at least two of the oligonucleotides hybridize to the template; and ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide. The method can further include a step of filling-in gaps. The method can further include a step of trimming flaps.

In another aspect, the invention is directed to a kit for performing a method of directed evolution comprising components for fragmenting oligonucleotides, for filling-in gaps, for trimming flaps, for proofreading, for incorporating uracils in templates and for modifying templates. The kit can further include instructions for performing a method of directed evolution.

In another embodiment, the invention is directed to a kit for performing a method of directed evolution comprising components consisting of lambda exonuclease and DNase I. The kit can further include additional instructions for performing a method of directed evolution. The kit can also include components such as those for filling-in gaps, for trimming flaps, for proofreading, for incorporating uracils in templates and for modifying templates.

In another embodiment, the invention is directed to a kit for performing a method of directed evolution comprising components consisting of lambda exonuclease, deoxyuracil triphosphate and uracil DNA glycosolase. The kit can further include instructions for performing a method of directed evolution. The kit can further include additional components such as those useful for filling-in gaps, for trimming flaps and for proofreading.

The present invention is also related, at least in part, to the discovery that mRNA is particularly well suited for forming the template for the RACHITT™

method. The utilization of an mRNA template allows for convenient production of templates and a higher affinity of binding when completed with DNA oligonucleotides.. Moreover, mRNA templates can be rapidly removed or degraded from the formed chimeric polynucleotides using reagents or enzymes specific for RNA. This is particularly advantageous when selective amplification of the chimeric polynucleotide is desired.

In one embodiment, the present invention is directed to a method for forming a chimeric polynucleotide including the steps of preparing a single-stranded template containing RNA; contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed. The template can be mRNA. In another embodiment, the method can include the step of selectively amplifying the chimeric polynucleotide relative to the mRNA template. This selective amplification can be achieved by an *in vitro* nucleic acid amplification.

The step of treating the hybridized oligonucleotides can be an enzymatic treatment. The population of oligonucleotides can include oligonucleotides generated by *in vitro* amplification or synthetic production. At least one oligonucleotide in the population of oligonucleotides can include a region of random sequence. Additionally, the single-stranded template can include at least one region of random or partially random sequence.

The method of the present invention can also include the additional step of selecting a chimeric polynucleotide for a particular characteristic.

The method of the present invention can also include additional steps such as trimming flaps, filling in gaps between hybridized oligonucleotides and/or ligating hybridized oligonucleotides. The step of trimming flaps can precede, follow, or occur concurrently with the step of filling in gaps between hybridized oligonucleotides. Gaps can be filled in using a polymerase with reverse transcriptase activity. Hybridized oligonucleotides can be ligated using ligases, such as *Taq* DNA ligase or T4 DNA ligase. The single-stranded mRNA template can be protected

during the step of trimming flaps, by annealing one or more pre-selected oligonucleotides to either or both ends of the template.

In another embodiment, the present invention is directed to a method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to at least one reference polynucleotide, including the steps of preparing a single-stranded polynucleotide template containing mRNA; contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; trimming flaps; filling in gaps between adjacently hybridized oligonucleotides; ligating immediately adjacently hybridized oligonucleotides to form at least one chimeric polynucleotide; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide. Further, the mRNA template can be protected during the trimming flaps step by annealing pre-selected oligonucleotides to either or both ends of the template.

In another aspect, the present invention is directed to a chimeric polynucleotide prepared according to the method of the invention.

In yet another aspect, the invention is directed to a method for preparing an mRNA transient template for use in forming a chimeric polynucleotide including the steps of preparing a DNA molecule comprising a suitable vector and an insert to be transcribed; transforming a suitable cell with the DNA molecule such that transcription of the DNA molecule occurs; lysing the transformed cell; and purifying the transcribed mRNA product from the cell lysate. The method of the present invention can include the step of protecting the mRNA template. The present invention is also directed to the template formed by the method described herein.

In another embodiment, the present invention is directed to a method for preparing an mRNA transient template *in vitro* for use in forming a chimeric polynucleotide including the steps of: preparing a DNA molecule comprising a suitable promoter and an insert to be transcribed; transcription of the DNA molecule utilizing elements required for transcription; and purifying the transcribed mRNA

product. The elements required for transcription can include a DNA-dependent RNA polymerase (*i.e.*, a polymerase with "transcriptase" activity). Additionally, the elements required for transcription can include wheat germ or reticulocyte extracts. In another embodiment, the method can include the step of protecting the mRNA template.

In another aspect, the present invention is directed to the template formed by the method described herein.

In a further embodiment, the method comprises preparing the single-stranded template. Two or more genes can be recombined in the process. The population of oligonucleotides can include cDNA or fragments thereof.

The present invention is also related, at least in part, to the discovery of modifications of single-stranded templates used to assemble oligonucleotides prior to recombination. In one embodiment, the present invention optimizes a combined trimming, polymerization and ligation step (hereinafter "TPL"). Such modifications to the single-stranded template lead to an increase in the degree of chimeragenesis in the products by limiting polymerization during the TPL step to filling in only short gaps between adjacently hybridized oligonucleotides. Since filling in gaps between oligonucleotides results in the synthesis of sequences complementary to the template, the overall degree of chimeragenesis of the eventual product is typically lowered when long gaps are filled in. Thus, limiting the polymerization step to short gaps can increase the overall degree of chimeragenesis of the product polynucleotide. Moreover, this improvement also increases the overall proportion of chimeric products formed by limiting the formation of non-chimeric products. Modifications to the single-stranded template include both physical and chemical modifications. Methods for making and using the modified polynucleotide templates are also provided.

In one embodiment, the present invention is directed to a method for forming at least one chimeric polynucleotide including the steps of preparing a single-stranded polynucleotide template; modifying the single-stranded polynucleotide template such that the degree of chimeragenesis is altered; contacting the modified template with a population of oligonucleotides, under conditions such that at least

two of the oligonucleotides hybridize to the template; filling in gaps between hybridized oligonucleotides on the template; and ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide. In a particular embodiment, the invention is directed to modifications of the single-stranded template or other components utilized by the method that affect the step of filling in gaps between hybridized oligonucleotides on the template. The filling in of gaps can be accomplished using a DNA polymerase such as T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase. A modification of a template can be either a chemical or physical modification. Such modifications include enzymatic modification of nucleotide residues or the annealing of short, *e.g.*, 6-100, *e.g.*, 6-20 bases, oligonucleotides to the single-stranded polynucleotide template. The invention can include additional steps of trimming flaps, amplification and selection of chimeric polynucleotides with specific properties.

In one embodiment, the present invention is directed to a method for preparing at least one chimeric polynucleotide. The method includes the steps of preparing a single-stranded polynucleotide template containing a plurality of uracil residues, contacting the template with a population of oligonucleotides in which at least two of the oligonucleotides hybridize to the template, treating the template with an enzyme, filling in gaps between hybridized oligonucleotides on the template, and ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide. The template can be treated with an enzyme such as a uracil glycosylase, such that the template is modified. The filling in of gaps can be performed by a DNA polymerase such as T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase. The method can utilize templates which contain between 0.01 and 45% uracil residues. The method can also include additional steps of trimming flaps, amplification and selection of chimeric polynucleotides with specific properties.

In another embodiment, a method for preparing a chimeric polynucleotide includes the steps of preparing a single-stranded polynucleotide template; hybridizing a pre-selected oligonucleotide to the template; contacting the template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; filling in gaps between hybridized

oligonucleotides on the template; and ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide. The pre-selected oligonucleotides can be such that they are resistant to nucleotide extension, *e.g.*, the oligonucleotides terminate in a dideoxy residue. Additionally, the pre-selected oligonucleotides can contain uracil residues. The method can further include a step of removing the pre-selected oligonucleotide and replacing it with an oligonucleotide that allows for nucleotide extension. The replacement of a uracil-containing oligonucleotide can be accomplished, for example, by treating the oligonucleotide with an enzyme such as UDG prior to nucleotide extension. This embodiment of the invention can also further comprise the steps of annealing a second pre-selected oligonucleotide immediately adjacent to the first pre-selected oligonucleotide, and allowing for nucleotide extension from this second oligonucleotide. The filling in of gaps can be performed by a DNA polymerase such as T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase. The method can also include the steps of trimming flaps, and selection and amplification of chimeric polynucleotides with specific properties.

In another embodiment, the invention is a method for preparing a modified single-stranded polynucleotide template suitable for use in forming a chimeric polynucleotide with an increased degree of chimeragenesis relative to a chimeric polynucleotide formed using an un-modified single-stranded polynucleotide template, including the steps of: obtaining a single-stranded polynucleotide by a method selected from the group consisting of isolating a polynucleotide from a suitable nucleic acid source, synthetically manufacturing a polynucleotide, cleaving the polynucleotide from a larger polynucleotide, and amplifying a polynucleotide obtained by any of these methods; and treating the single-stranded template such that the template is modified an a manner that the step of filling in gaps is altered, thereby increasing the degree of chimeragenesis of the polynucleotide product obtained as a result of the present invention. The modification can be a chemical modification, *e.g.*, one that creates an abasic residue in the polynucleotide template, or a physical modification, *e.g.*, annealing a pre-selected oligonucleotide to the polynucleotide template.

The present invention is related, at least in part, to the discovery that the single-stranded polynucleotide template can be prepared *in vivo* using a vector derived from a bacteriophage genome that contains an inserted polynucleotide of interest. In a particular embodiment, the bacteriophage is M13 which, as a mature phage particle, contains a single-stranded DNA copy of its genome. Utilizing M13-derived vectors, inserts can be amplified and isolated as single-stranded polynucleotides contained within the M13-derived vector, thus abrogating the need to eliminate the complementary strand of a double-stranded molecule present when polynucleotides are prepared via a different method. Oligonucleotides can assemble onto sequences corresponding to the polynucleotide insert contained in the M13-derived vector. In addition, depending on the host bacterial strain and conditions of growth used to amplify the M13 phage, the template can be synthesized so as to facilitate the selective amplification of the chimeric polynucleotide product relative to the template.

In one embodiment, the present invention is directed to a method for forming a chimeric polynucleotide including the steps of: contacting a single-stranded template comprising a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed. The phage vector can be derived from bacteriophage M13. Additionally, the single-stranded template can be isolated from a biological source. The bacteriophage can be amplified in a bacterial strain defective in one or more dUTPases and uracil N-glycosylases, such as a bacterial strain containing a genome with a *dut⁻/ung⁻* genotype. Further, the single-stranded template can contain at least one uracil residue.

Treating the hybridized oligonucleotides can be accomplished by an enzymatic treatment.

The template and at least one parent polynucleotide used to form the population of oligonucleotides can encode a desulfurizing enzyme or fragment thereof.

The population of oligonucleotides can include oligonucleotides generated by *in vitro* amplification or synthetic production. The population of oligonucleotides can also include oligonucleotides with a region of random or partially random sequence. Additionally, the single-stranded template can include at least one region of random or partially random sequence.

This method can be repeated using one or more chimeric polynucleotides to generate the template or the population of oligonucleotides.

In another embodiment, the further steps of trimming flaps, filling in gaps between hybridized oligonucleotides and/or ligating hybridized oligonucleotides can be included in the method of the present invention. The step of trimming flaps can precede, follow, or occur concurrently with the step of filling in gaps between adjacently hybridized oligonucleotides. Gaps can be filled in using a polymerase such as T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase. Hybridized oligonucleotides can be ligated by using appropriate ligases such as *Taq* DNA ligase or T4 DNA ligase.

In another embodiment, the present invention is directed to a method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule are altered in comparison to at least one reference polynucleotide, including the steps of: contacting a single-stranded template comprising a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; trimming flaps; filling in gaps between hybridized oligonucleotides; ligating hybridized oligonucleotides to form at least one chimeric polynucleotide; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide. The single-stranded template can be derived from bacteriophage M13. The single-stranded template can contain at least one uracil residue. The treating of the hybridized oligonucleotides can be accomplished by an enzymatic treatment. The template and at least one parent polynucleotide used to form the population of oligonucleotides can encode a desulfurizing enzyme or

fragment thereof. The population of oligonucleotides can include oligonucleotides generated by *in vitro* amplification or synthetic production. The population of oligonucleotides can include at least one oligonucleotide with a region of random or partially random sequence. The single-stranded template can include at least one region of random or partially random sequence. The method of this embodiment can be repeated with one or more chimeric polynucleotides used to generate the template or population of oligonucleotides.

In another embodiment, the present invention is directed to a method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule are altered in comparison to at least one reference polynucleotide, including the steps of: contacting the single-stranded template including of a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; treating the hybridized oligonucleotides such that a heteroduplex polynucleotide containing a contiguous chimeric polynucleotide and the single-stranded template is formed; and transforming a bacterial strain with the heteroduplex. The single-stranded template can be derived from bacteriophage M13. This method can further include the steps of trimming flaps, filling in gaps or ligating immediately adjacently hybridized oligonucleotides. This method can also further include the steps of selectively amplifying the chimeric polynucleotide over the single-stranded template; and selecting chimeric polynucleotides with optimized properties when compared to the reference polynucleotide.

In yet another embodiment, the invention is directed to a method for the production of a library of chimeric polynucleotides including the steps of: preparing a plurality of single-stranded templates that have a phage vector and a polynucleotide insert; contacting the single-stranded templates with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to more than one template; and treating the oligonucleotides hybridized to each template such that more than one contiguous chimeric polynucleotide is formed, thereby generating the library of chimeric polynucleotides. In another embodiment,

the invention is directed to a library of chimeric polynucleotides produced by the method described herein.

In yet another embodiment, the invention is directed to a method for producing a single-stranded polynucleotide template for use in forming a chimeric polynucleotide including the steps of: inserting a polynucleotide into a phage vector; transforming bacterial strain with the vector containing the insert; inducing phage growth; harvesting phage particles; and isolating single-stranded DNA from the phage particles; thereby resulting in the recovery of the single-stranded polynucleotide template. In another embodiment, the present invention is directed to the single-stranded polynucleotide template produced according to the method described herein.

A method for forming a chemically modified single-stranded polynucleotide template for use in a method of directed evolution comprising preparing a double-stranded polynucleotide comprising the single-stranded polynucleotide template and a complementary polynucleotide strand; denaturing the double-stranded polynucleotide;

adding a single-stranded oligonucleotide capable of annealing to the strand complementary to the single-stranded template; and isolating the single-stranded polynucleotide template from its complementary strand and from the added oligonucleotide, thus yielding the purified single-stranded polynucleotide template.

A method for generating a chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide comprising the steps of preparing a double-stranded polynucleotide comprising a single-stranded polynucleotide template and a complementary polynucleotide strand, denaturing the double-stranded polynucleotide; adding a single-stranded oligonucleotide capable of annealing to the strand complementary to the single-stranded template; isolating the single-stranded polynucleotide template from its complementary strand and from the added oligonucleotide, thus yielding the purified single-stranded polynucleotide template; contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; trimming flaps; filling in gaps;

ligating hybridized oligonucleotides to form at least one chimeric polynucleotide; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and selecting or screening the chimeric polynucleotide, wherein a characteristic is altered in comparison to the reference polynucleotide.

A method for forming a chemically modified single-stranded polynucleotide template for use in a method of directed evolution comprising preparing a double-stranded polynucleotide comprising the chemically modified single-stranded polynucleotide template and a complementary polynucleotide strand; denaturing the double-stranded polynucleotide; and isolating the single-stranded polynucleotide template from its complementary strand, thus yielding the purified single-stranded polynucleotide template.

A method for generating a chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide, comprising the steps of contacting the single-stranded template comprising a chemically modified single-stranded polynucleotide, which modification allows the single-stranded template to be immobilized on a solid matrix, and sequences derived from parent polynucleotides, with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and selecting or screening the chimeric polynucleotide, wherein a characteristic is altered in comparison to the reference polynucleotide.

A method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to at least one reference polynucleotide, comprising the steps of contacting a single-stranded template comprising sequences derived from at least one parent polynucleotide and at least one additional sequence whereby the single-stranded template is about 25% larger than a parent polynucleotide, with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide.

In one embodiment, the invention is directed to method for forming a chimeric polynucleotide including preparing a random population of oligonucleotides from at least one nucleic acid with a preselected nucleotide sequence, contacting a single-stranded template with the population of oligonucleotides under conditions such that at least two of the oligonucleotides hybridize to the template, and ligating the hybridized oligonucleotides such that a chimeric polynucleotide is formed. The method can further include a step of filling in gaps. In particular embodiments, the oligonucletides comprise lengths in the range from about 10 nucleotides to about 200 nucleotides. In a particular embodiment, the hybridization conditions have been optimized to promote the annealing of short oligonucleotides. The treatment to form the chimeric polynucleotide can utilize a cryophilic enzyme.

In another embodiment, the invention is directed to a method for producing a single-stranded polynucleotide template for use in forming a chimeric polynucleotide including the steps of transcribing a double-stranded polynucleotide comprising a phage promoter to create an RNA transcript, reverse transcribing the RNA transcript to create a hybrid polynucleotide, and degrading the RNA strand of the hybrid polynucleotide, thus resulting in a single-stranded polynucleotide template for use in forming a chimeric polynucleotide.

In another aspect, the invention is direct to the chimeric polynucleotide formed.

In another embodiment, the invention is directed to a method of forming a randomly fragmented population oligonucleotides for use in forming at least one chimeric polynucleotide including treating a double-stranded polynucleotide comprising a parent polynucleotide such that a plurality of modified bases are formed on either or both strand, and treating the double-stranded polynucleotide such that single-strand nicks are created as a result of the treatment and the modified bases, thus forming a randomly fragmented population of oligonucleotides for use in forming at least one chimeric polynucleotide.

In another aspect, the invention is direct to the chimeric polynucleotide formed.

In one embodiment, the invention is directed to a method of forming a population of single-stranded oligonucleotides for use in forming at least one chimeric polynucleotide including obtaining at least two double-stranded polynucleotides, treating the double-stranded polynucleotides with an exonuclease, thereby obtaining from each double-stranded polynucleotide a strand containing a 3' phosphate, annealing the 3' phosphate strands to form heteroduplex nucleic acids, treating the heteroduplex nucleic acids with an enzyme that cleaves mismatches to yield homoduplexes, and treating the homoduplexes with an enzyme to degrade the strand containing the incorporated uracils, thereby forming a population of single-stranded oligonucleotides for use in directed evolution. In a particular embodiment, the enzyme that cleaves mismatches is a branch resolving enzyme. The branch resolving enzyme can be T4 endonuclease VII or T7 endonuclease I. The method can further include fractionation of the single-stranded oligonucleotides based on size. The first and second double-stranded nucleic acid can be from a single source.

In another aspect, the invention is directed to the population of single-stranded oligonucleotides produced according to the method.

In another embodiment, the invention is directed to a method of forming a population of oligonucleotides for use in forming at least one chimeric polynucleotide including obtaining a double-stranded nucleic acid polynucleotide,. obtaining a second double-stranded nucleic acid, denaturing both double-stranded nucleic acids thereby obtaining single-stranded polynucleotides, annealing the single-stranded polynucleotides to form heteroduplex nucleic acids, and treating the heteroduplex nucleic acids with an enzyme that cleaves mismatches to yield oligonucleotides, thereby forming a population of oligonucleotides for use in forming at least one chimeric polynucleotide.

In another aspect, the invention is directed to the population of oligonucleotides formed.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram illustrating an embodiment of a method of the present invention.

Figure 2 is a schematic diagram showing an example of an oligonucleotide transient template heteroduplex.

Figures 3A, 3B, 3C and 3D set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszA* from *Rhodococcus sp.* strain IGTS8 (SEQ ID NOS: 8 and 9, respectively).

Figures 4A, 4B and 4C set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszB* from *Rhodococcus sp.* strain IGTS8 (SEQ ID NOS: 10 and 11, respectively).

Figures 5A, 5B and 5C set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszC* from *Rhodococcus sp.* strain IGTS8 (SEQ ID NOS: 12 and 13, respectively).

Figures 6A, 6B, 6C and 6D set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszA* from *Sphingomonas sp.* strain AD109 (SEQ ID NOS: 14 and 15, respectively).

Figures 7A, 7B and 7C set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszB* from *Sphingomonas sp.* strain AD109 (SEQ ID NOS: 16 and 17, respectively).

Figures 8A, 8B and 8C set forth the DNA sequence and predicted amino acid sequence of the desulfurization gene *dszC* from *Sphingomonas sp.* strain AD109 (SEQ ID NOS: 18 and 19, respectively).

Figure 9 shows the results of a microtiter-format assay for DBT monooxygenase activity of 7 chimeric clones obtained by the method of the present invention.

Figure 10 shows a DNA sequence alignment of a *dszC* chimera of the present invention (SEQ ID NO: 3) with the parental sequences of *dszC* from *Nocardia sp.* strain A3H1 (SEQ ID NO: 23) and *Rhodococcus sp.* strain IGTS8 (SEQ ID NO: 24).

Figure 11 is an illustration of the life cycle of a representative virulent bacteriophages.

Figure 12 depicts an embodiment of the present invention using an immobilized single-stranded template.

Figure 13 is a schematic of various arrangements of the flanking and parent sequences in the single-stranded transient template.

Figure 14 is a schematic representation of one arrangement of a transient template with a plurality of parent-derived sequences.

Figure 15 is a schematic representation of a method of generating donor fragments utilizing a structure-specific endonuclease.

Figures 16A and 16B are schematic representations of a method of using pre-selected oligonucleotides which hybridize to either end of the template.

DETAILED DESCRIPTION OF THE INVENTION

The method of the present invention facilitates the generation of chimeric polynucleotides. "Chimeric polynucleotides," as used herein, contain nucleotide sequences from multiple sources, typically related sequences or otherwise similar polynucleotides, referred to herein as "parent polynucleotides". Any number of parent polynucleotides can be utilized. In particular embodiments two, three and four parents are present. The partial sequences are usually presented in the form of an "oligonucleotide population". The present method facilitates the recombination of regions of parent polynucleotides derived or isolated from any suitable source of nucleic acid, including known sequences, unknown or uncharacterized sequences, *e.g.*, nucleic acid molecules isolated from environmental or other sources, and regions of randomly generated sequence, into a chimeric polynucleotide. In addition, the present method allows for deletion and/or insertion of sequences from reference parent polynucleotides compared to the resultant chimeric polynucleotides.

The RACHITT™ method results in low levels of siblings (additional clones produced by replication rather than separate recombination events), inactive proteins or unshuffled parental clones. The method also yields an unprecedented

predominance of highly chimeragenic (mosaic) clones and provides the unique ability to effect frequent recombination between close and even adjacent alleles. These observed improvements are in the key mechanisms that determine the diversity and evolutionary potential of gene shuffled libraries. These elements combined indicate an ability to explore more diverse permutations of multiple mutations with the RACHITT™ method than has been possible using other methods. The ability to capture untapped regions of sequence space better allows the evolution of polynucleotides encoding proteins having desired characteristics.

While not wishing to be bound by theory, the advantages of the RACHITT™ method can be attributed, at least in part, to the aspects of the method as described below. The RACHITT™ method utilizes a single round of hybridization, as opposed to reiterative thermocycling, and thus lends itself to greater specificity in similar or more divergent pairings. When necessary, it also incorporates "flap" trimming to result in perfectly duplexed ligatable/extendable ends from otherwise unproductive fragments. Trimming also allows incorporation of fragments much smaller that those generated by DNase I. In addition, many DNA ligases allow ligation of mismatched termini. Where flaps are not perfectly trimmed, such low-fidelity ligation may contribute to incorporation of fragments with adjacent mismatches. Ligation of adjacent fragments from the same parent also allows linking of larger stretches from any given parent than were present after DNase I cleavage. Finally, the RACHITT™ method typically exploits a bottom strand template from one parent and top strand fragments from the same and/or other parents. This prevents parental fragments from reannealing to their own complementary strands.

In one embodiment, the invention is drawn to a method of forming at least one chimeric polynucleotide including the steps of contacting at least one single-stranded polynucleotide template with a population of oligonucleotides, under conditions wherein at least two oligonucleotides hybridize to the template and treating the hybridized oligonucleotides to form at least one chimeric polynucleotide.

The methods described herein comprise process steps involved in the formation of chimeric polynucleotides. Reference is now made to Figure 1. Figure 1 depicts schematically the steps utilized by one embodiment of the present invention

in forming a chimeric polynucleotide. First, fragments of a population of related oligonucleotides are generated 10 by, for example, random cleavage or partial DNA synthesis or degradation. These oligonucleotide fragments are assembled onto a single-stranded template 20 in such a way as to allow for fragments derived from different molecules to assemble onto the same template. In some cases, overlaps occur, thus creating "flaps" 30. The term "flaps" is intended to include the unhybridized portion of an oligonucleotide that is otherwise hybridized to a template. In other cases, regions of the template remain single-stranded, thus creating "gaps" 40. Flaps are trimmed and gaps are filled in prior to the generation of a contiguous chimeric polynucleotide by ligating the assembled oligonucleotides 50. The product chimeric oligonucleotide can then be selectively cloned, amplified, selected or screened with respect to the template 60 for optimized traits of interest 70. Although a wide variety of templates are within the scope of the present invention, typically, the template is selected such that it contains "target" sequences that are related to the parent polynucleotides that are used to generate the population of oligonucleotides. Target sequences are related to the parent polynucleotides such that they allow for annealing of parent-derived oligonucleotides. As such, the template allows for the assembly of parent-derived oligonucleotides which, following the completion of the steps of the methods of the invention, result in the formation of a chimeric polynucleotide product.

The template of the present invention can be formed from combinations of any molecules which allow for sequence specific annealing of oligonucleotides. Examples of such molecules include nucleic acids such as DNA, *e.g.*, cDNA, RNA, *e.g.*, mRNA, and other molecules such as "Peptide Nucleic Acids" (hereinafter "PNA"). The term PNA as used herein includes polymers, *e.g.*, naturally occurring polyamides (pseudopeptides) which can hybridize to nucleic acids, *e.g.*, DNA and RNA, with sequence specificity (see United States Patent Nos.: 5,539,082; 5,527,675; 5,623,049; 5,714,331; 5,736,336; 5,773,571; 5,786,461; and Egholm *et al.*, *Nature 365:* 566-568 (1993), the teachings of which are incorporated by reference herein in their entireties). Such molecules can comprise natural bases, *e.g.*, adenine, thymine, cytosine, guanine and uracil, analog bases, *e.g.*, inosine,

bromouracil and nitroindole, or modified bases *e.g.*, biotin labeled or a digoxygenin labeled or a combination thereof.

A template of the invention can be of any length. The template can be from about 10 to about 100,000 nucleotides in length. In a particular embodiment of the present invention, the template can be from about 100 to about 10,000 nucleotides in length. In a more particular embodiment, the template can be from about 200 to about 1000 nucleotides in length. The template can also be about the length of a gene of interest, such genes can be in the form of a genomic copy or a cDNA copy, *e.g.*, without introns. The template can comprise the entire coding sequence or a portion of the coding sequence of a gene of interest. In particular embodiments, the template comprises a significant portion of the coding sequence of a gene, *e.g.*, 50%, 60%, 70%, 80%, 90% or 95% of the coding sequence of a gene. In particular embodiments, regulatory sequences can also be included in the template. Templates of a shorter length, for example, from about 10 to 250 nucleotides in length are also considered to be within the scope of the invention and are useful in particular embodiments.

The template can be produced *in vitro* using enzymatic or chemical means. Methods of *in vitro* production of nucleic acid sequences are well known in the art. The template can be circular or linear. The template is preferably single-stranded. The template can comprise sequences in addition to target sequences, thus creating a template which is larger than the parent polynucleotides. For example, the template can be nicked or gapped, wherein "gapped" refers to a partially double-stranded molecule comprising a target sequence and regions flanking the target sequence, wherein all or a portion of the target sequence is single stranded.

The template can include or preserve one or more regions with functional characteristics or structural motifs of the parent polynucleotides. These regions can include nucleic acid structural motifs, protein binding domains, metal binding domains, nucleic acid binding domains, domains with enzymatic activity, or fragments of these domains. These regions can include ribozymes, deoxyribozymes, promoters, enhancers, origins of replication, open reading frames, or fragments thereof. These regions can encode aptamers, wherein aptamers are small single- or

double-stranded DNA or RNA molecules that bind specific molecular targets (Bock *et al.*, *Nature 355*:564-566, 1992; Ellington and Szostak, *Nature 346*:818-822, 1990; and, Werstuck and Green, *Science 282*:296-298, 1998, the teachings of which are hereby incorporated by reference in their entireties).

The template of the present invention can also include sequences that are not known to have any particular function. These sequences can be selected from any known source of nucleic acid, including sequences synthesized *in vitro*; or these sequences can be random or partially random. Random or partially random sequences can be generated by synthesizing a polynucleotide based on a known sequence, except that a portion of the sequence is randomized, *e.g.*, randomizing the last 50 nucleotides, or wherein certain positions within the sequence are randomized, *e.g.*, randomizing particular codon(s) of a coding sequence or wherein certain bases are randomized, *e.g.*, randomizing all adenines. These sequences can further encode proteins or domains of proteins including folding structures or structural motifs; binding domains such as protein binding domains, metal binding domains, co-factor binding domains, lipid binding domains and nucleic acid binding domains; domains with enzymatic function; sites for allosteric or competitive inhibition and the like; or fragments of these domains. These regions can also include amino acid sequences that are not known to have any particular function or can be randomized amino acid sequence.

In a particular embodiment, the present invention utilizes an RNA template, prepared by methods known to those of skill in the art, which allows for the selective amplification of the chimeric polynucleotide over the mRNA template. The utilization of an RNA template allows for the convenient production of single-stranded polynucleotide template molecules without a double-stranded intermediate.

RNA, preferably, mRNA, can be utilized as a transient template in certain embodiments during the RACHITT™ method. As used herein, the term "transient template" is intended to reflect that the template is not directly incorporated into the resulting chimeric molecules. An mRNA template can be prepared using methods known to those of skill in the art (see *Molecular Cloning.- A Laboratory Manual, second edition*, edited by Sambrook, Fritsch, & Maniatis. Cold Spring Harbor

Laboratory (1989), 18.81-18.88, the teachings of which are hereby incorporated by reference in its entirety). For example, mRNA can be synthesized by *in vitro* transcription reactions utilizing commercially available bacteriophage DNA-dependent RNA polymerases (*e.g.*, those derived from T3 or T7), or cell extracts containing all components necessary for transcription (*e.g.*, wheat germ extracts or reticulocyte extracts). Alternatively, the template can be derived from a larger molecule, *e.g.*, the template can be a fragment of an mRNA molecule.

Subsequent to transcription, modifications to the transcript can occur such as adding the 5' cap, polyadenylation (creating a "poly-A tail") and splicing. In addition to altering the coding region of the transcript, as occurs with splicing, post-transcriptional modifications can affect the half-life of the transcript. RNA is inherently less stable than DNA and this can be alleviated at least in part by adding the 5' cap and poly-A tail. Indeed, the instability of RNA has led to the development of artificial ways of increasing the half-life for the template such as working in the presence of diethylpyrocarbonate (DEPC) which inhibits RNases, attaching chemical adducts to the ends of the RNA, and making the ends of the RNA double-stranded by adding short DNA oligonucleotides. Although substantial post-transcriptional modification occurs for many mRNA transcripts, an mRNA transcript can be utilized as a template for the present invention either before or after post-transcriptional modification (*e.g.*, generation of the 5' cap, polyadenylation, methylation, and splicing). The term "mRNA" as used herein, is not intended to be limiting as to the specific state of modification of the transcript.

In one embodiment, the present invention is directed to a method for forming a chimeric polynucleotide including the steps of preparing a single-stranded template containing RNA; contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed.

In addition to *in vitro* synthesis for preparing the RNA template, the template can be isolated from an organism, such as an eubacterial, archeal, eukaryotic or viral organism.

The template can also be formed using an mRNA intermediate. For example, a double-stranded DNA molecule comprising a transcriptional promoter can be transcribed to generate a single-stranded mRNA molecule. This mRNA molecule can then be reverse transcribed using a RNA-dependent DNA polymerase (a "reverse transcriptase"), resulting in a RNA:DNA hybrid. Such hybrids can be treated with nucleases such as RNase H, which selectively degrade the RNA strand, thus resulting in a single-stranded DNA molecule which is an exact copy of one of the starting strands. If the starting double-stranded DNA molecule contained target sequences, then the result of transcription followed by reverse transcription and RNA digestion is a single-stranded template suitable for use in the RACHITT™ method.

A template can be modified to regulate the degree of chimeragenesis of the formed chimeric polynucleotide. Such modification will also increase the overall proportion of chimeric products formed. These modifications can be physical or chemical. An example of a physical modification is modifying the template by annealing a specific oligonucleotide to the template. An example of a chemical modification is covalently attaching a specific chemical group to nucleotide residues contained in the polynucleotide.

In a particular embodiment, the present invention utilizes a single-stranded polynucleotide template isolated from a bacteriophage. In a preferred embodiment, an M13 bacteriophage is utilized. M13 bacteriophage particles contain the bacteriophage genome as a single-stranded DNA molecule that can be used as a "vector." The term "vector" is intended to include any DNA molecule that can contain and replicate an exogenous DNA polynucleotide of interest, and that can be amplified and isolated by methods commonly known to those of ordinary skill in the art.

Figure 11 depicts a typical bacteriophage infection event into a host bacterium 90. Virulent filamentous bacteriophages, *e.g.*, M13, exhibit a life cycle divided into stages of phage attachment to the host cell surface 100, insertion of phage genome into the host cell 110, particle production and DNA replication 120, packaging 130 and lysis 140. Adsorption onto the host cell surface and subsequent entry of the phage genome into the host cell is sometimes mediated by the F-encoded

pilus. In the specific case of M13, a single-stranded DNA molecule is inserted and the host machinery is recruited to synthesize the complementary strand. Genes are transcribed to promote phage head assembly as well as other late life cycle events. Eventually, the DNA circle is replicated, typically by a rolling circle mechanism, which produces many copies of the single-stranded genome. These copies are processed and packaged into the newly assembled phage heads. Eventually, the host cell is lysed and the new phage particles are released to repeat the cycle in other hosts. The phage life cycle, then, provides a mechanism for the amplification of a single-stranded DNA molecule.

In one embodiment, the present invention is a method for forming a chimeric polynucleotide including the steps of preparing a single-stranded template containing bacteriophage DNA with an insert; contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed. The step of treating the hybridized oligonucleotides can be an enzymatic treatment. The population of oligonucleotides includes oligonucleotides generated by *in vitro* or *in vivo* amplification or synthetic production. At least one oligonucleotide in the population of oligonucleotides can include a region of random or partially random sequence. Additionally, the single-stranded template can include at least one region of random or partially random sequence.

In a particular embodiment, the method utilizes bacteriophage M13 vector DNA (for review, see Messing, J. *Methods Enzymol.* 1983. 101:20-78) containing a polynucleotide of interest as a transient template during the RACHITT™ method. The single-stranded template can be prepared using many methods known to those of skill in the art. The nucleic acid for use as a template can be synthetically manufactured or isolated from any suitable source of nucleic acid. In a particular embodiment for producing a template for use in the methods of the invention, a polynucleotide of interest can be cloned into bacteriophage M13 DNA and amplified by infecting a suitable host bacterial strain. Methods for phage harvest and lysis are commonly known in the art as are methods for DNA isolation from phage lysates

(*Current Protocols in Molecular Biology*. "Preparing and Using M13-Derived Vectors." John Wiley and Sons, New York, section 1.15.1). In addition, the template can be derived from a larger molecule, *e.g.*, the template can be a fragment of an isolated bacteriophage genome.

M13 is a bacteriophage capable of infecting, for example, *Escherichia coli* (*E. coli*). The mature phage particle contains a circular single-stranded DNA copy of the phage genome. The insertion of convenient restriction sites into the phage genome allows for the use of these phage-derived vectors in the production of single-stranded copies of specific polynucleotide sequences. Such M13-derived vectors with convenient restriction sites are commercially available (*e.g.*, M13mp18 and M13mp19; New England BioLabs, Beverly, MA). The present invention utilizes parent-derived polynucleotide sequences inserted into M13-derived vectors. Single-stranded templates, amplified as part of a phage-derived vector, avoid double-stranded *in vitro* DNA intermediates required by other methods of preparation, and, thus, do not require a step of digesting one polynucleotide strand of a double-stranded molecule in order to produce a single-stranded template. Using an M13-derived vector, a polynucleotide sequence of interest can be inserted into the vector and a single-stranded copy can be isolated for use in subsequent directed evolution.

Additionally, when M13 infects an *E. coli* strain that is defective in dUTPase and uracil N-glycosylase activity (hereinafter referred to as "*dut⁻/ung⁻*" strains), phage particles containing circular single-stranded DNA molecules containing uracil residues can be isolated. Oligonucleotides can be annealed to these uracil-containing molecules during the RACHITT™ method, and the resultant partially or fully double-stranded molecules can be transformed into an appropriate host cell which has dUTPase and uracil N-glycosylase activities. The host cell then completes the synthesis of the second strand on partially double-stranded molecules with the previously annealed oligonucleotides incorporated into the newly synthesized strand. Alternatively, the uracil containing strand is not replicated and a newly synthesized strand is made. During host cell replication and division, the host replication and repair machinery preferentially corrects mismatches between strands by selecting the strand without uracils as the correct strand. Thus, the newly synthesized stand

containing hybridized RACHITT™ oligonucleotides is replicated and the original single stranded template is lost to degradation and dilution. This preference has traditionally been exploited for creating site directed mutants of inserts in M13 vectors (Kunkel, T.A., *Proc. Natl. Acad. Sci.* USA 1985. 82:488-492, the teachings of which are hereby incorporated by reference in its entirety).

The template can be derived from a double-stranded DNA molecule. Many methods are known in the art for isolating a single strand of a double-stranded DNA molecule. For example, a double-stranded molecule can be generated by restriction endonuclease digestion or by PCR with a primer such that one strand is protected from exonucleases. Treating the double-stranded product with an exonuclease, such as bacteriophage λ exonuclease, specifically eliminates the non-protected (*i.e.*, the non-template) strand.

Alternatively, the single-stranded template can be separated and purified away from its complementary strand. A double-stranded molecule obtained, for instance, as the result of an *in vitro* amplification of parent polynucleotides or through isolation from natural sources, can be denatured to separate the template strand from the non-template strand. The present invention describes methods for specific removal of the non- template strand. For example, after denaturing the template from its complementary strand, one or more oligonucleotides can be annealed to the non-template strand, thus rendering this complex containing the non-template strand both larger and partially double-stranded. The template and non-template strands are now distinguishable by techniques known in the art such as electrophoresis, chromatography and gel filtration. The present invention is further directed to preparing a single-stranded template that is chemically modified at one end. The modification is such that it allows for the specific immobilization of the template strand rather than immobilization of the non-template strand. For example, a double-stranded polynucleotide can be generated using a primer pair to a parent polynucleotide in which one of the primers is biotinylated. The result is a molecule in which one strand is biotinylated and the other is not. The two strands can be separated after denaturation if they are exposed to a streptavidin-agarose matrix. Streptavidin binds biotin, and, thus, binds all polynucleotide strands which are

biotinylated. The non-template strand complementary to the single-stranded template can be simply washed away.

A general overview of this approach is diagramed in Figure 12. This Figure shows the contacting of a double-stranded polynucleotide 190, one strand of which has been chemically modified (denoted by an asterisk), with a solid matrix containing an affinity moiety specific to the chemical group which is used to modify the template strand 200. Attachment of the polynucleotide to the matrix 210 followed by denaturation and washing, yields a single-stranded template attached to a solid matrix 220. This can be used as a substrate for the RACHITT™ method. After a chimeric polynucleotide has been formed as a result of the RACHITT™ method 230, the transient template can be separated from the chimeric RACHITT™ product by denaturation and washing 240. In this case, the chimeric product is isolated and specifically amplified with respect to the single-stranded transient template. This method for preparing a single-stranded transient template can be automated, and also advantageously requires fewer enzymatic steps than other methods of preparation, improves template stability, generates potentially re-usable templates, and allows for both easy purification of the single-stranded template and the subsequent specific amplification of the chimeric RACHITT™ product over the transient template.

In particular embodiments, the single-stranded template also contains sequences ("flanking" sequences) derived from sources other than parent polynucleotides, thus making the templates longer than the parent polynucleotide. The arrangement of parent, or "target," and flanking sequences is diagramed in Figure 13. Target sequences 290 are made longer by adding flanking sequences 300 to the 5' end 310, the 3' end 320 or to both ends 330 of the parent-derived sequence. Use of a larger template allows for easier purification of the single-stranded template and for the selective amplification of the chimeric polynucleotide product.

Figure 14 depicts another arrangement of the transient template. In this Figure, two target sequences 360 and 370 are separated by a nucleotide sequence that does not bind the oligonucleotide fragments 380, *i.e.*, a "linker" sequence. The two target sequences can be derived from the same or different parent polynucleotides. An advantage of using such a template is that, after the two chimeric polynucleotides

are formed 390 and 400, both ends of the template/chimera hybrid are double-stranded, and, thus, resistant to exonuclease activity.

In one embodiment, the present invention is directed to a method for forming a chimeric polynucleotide including the steps of contacting a single-stranded template containing sequences derived from a parent polynucleotide and at least one additional sequence whereby the single-stranded template is at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, or 90% larger than the parent polynucleotide, with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template generating a chimeric strand; selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide.

The template of the present invention is derived from at least one parent polynucleotide. The template of the present invention also contains sequences in addition to the parent-derived sequences. These additional sequences increase the length of the template, in a preferred embodiment, by at least about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, or 90%. Such a template can be isolated or synthesized by methods known in the art. For example, single-stranded DNA templates can be synthesized *in vitro* in enzymatic amplification reactions or using automated DNA synthesizers. In addition to *in vitro* synthesis of the single-stranded transient template, parent polynucleotides can be isolated from an organism, such as a eubacterial, archeal, eukaryotic or viral organism.

The template of the present invention can contain elements that allow for the selective amplification of the resultant chimeric polynucleotides over the templates. For example, the additional sequences present in the template of the present invention result in the formation, after the RACHITT™ method is employed, of a partially double-stranded molecule - one strand being the template and the other strand being the chimeric polynucleotide. Since the chimeric polynucleotide was formed based on parent-derived sequences, the template strand will be about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, or 90% longer. After

denaturation, the chimeric product can be separated from the template by methods designed to separate molecules based on size (*e.g.*, electrophoresis and size-exclusion chromatography), charge (*e.g.*, ion-exchange chromatography), or sequence (*e.g.*, DNA affinity chromatography). Moreover, separation can be accomplished by synthesizing the single-stranded template such that it contains an affinity moiety that can be selectively bound or eluted during affinity chromatography. Examples of such moieties are biotin and digoxygenin, which can be affinity separated along with the molecule to which each is attached, using avidin or anti-digoxygenin antibodies, respectively.

A template of the invention is contacted with a population of oligonucleotides to form a chimeric polynucleotide. The oligonucleotides of the present invention comprise molecules that hybridize to the templates of the invention, *e.g.*, nucleic acids, such that the oligonucleotide sequences can ultimately be joined to form contiguous chimeric molecules.

Typically, the population of the oligonucleotides comprises a random population. The term "random" as used herein is intended to reflect an absence of preselection. Such absence can be of any degree; it need not be a total absence of preselection, nor does the term indicate a requirement for an absence of preference or bias. The term can be used to describe populations of oligonucleotides, sequences, events, processes, states or conditions, or other such terms. Such compositions can range over a span of values and any one component can occupy any of these values. For example, a population of oligonucleotides which is generated by the digestion of two genes with a restriction enzyme is a "random population" when the particular oligonucleotides formed by the process are not preselected, for example, during a partial digestion. This is true even when the gene sequences are known and the restriction enzyme preferentially cleaves a particular site. Sequences can be random if at least one position in the sequence is not specifically defined (for example, if at least one position of an oligonucleotide could be and is either one of two or more nucleotides). The randomly fragmented population of oligonucleotides can include oligonucleotides wherein a portion of the oligonucleotides comprise random or partially random sequence as described herein.

As described herein, the oligonucleotides can include both DNA and RNA and modifications thereof. The DNA or RNA can comprise natural bases, *e.g.*, adenine, thymine, cytosine, guanine and uracil, analog bases, *e.g.*, inosine, bromouracil and nitroindole, modified bases, *e.g.*, 8-oxo-dGTP, biotin labeled or digoxygen in labeled base, or any combination thereof. The oligonucleotides can comprise single or double-stranded molecules. Double-stranded oligonucleotides can be made single-stranded, for example, by denaturing prior to hybridization. Methods of denaturing and annealing oligonucleotide sequences, including peptide nucleic acid sequences, are well known in the art.

Oligonucleotides can be single-stranded and of opposite orientation when compared to the orientation of the template. In particular embodiments, the oligonucleotides of the population do not hybridize to each other, *e.g.*, do not self-anneal. However, it is clear that one of skill in the art can skew the availability of any given oligonucleotide to hybridize to a template by including an oligonucleotide capable of hybridizing to the oligonucleotide.

Oligonucleotides can be of any length, for example, from about 5 to about 50,000 nucleotides in length. In a particular embodiment, oligonucleotides can be from about 16 to about 400 nucleotides in length. In a still more particular embodiment, the oligonucleotides can be from about 20 to about 200 nucleotides in length. Additionally, in certain embodiments, oligonucleotides can be quite small, ranging in size, for example, from about 5 to 10 nucleotides; preferably ranging from about 10 to 50 nucleotides in length.

In a particular embodiment, the population of oligonucleotides can include DNA that is reverse transcribed from RNA (hereinafter known as "cDNA"). Such cDNA can comprise all or a portion of at least one cDNA library. A "cDNA library" is used herein to describe a population of DNA molecules in which the DNA molecules represent all or a portion of at least two different genes. In particular embodiments, the population of oligonucleotides can contain nucleic acid from at least two, three or four different genes which can be recombined to form a chimeric molecule.

A population of oligonucleotides can comprises molecules generated using any one of the methods described herein or any combinations thereof. A random population of oligonucleotides can be generated by randomly fragmenting, digesting or elongating at least one polynucleotide of interest. A random population of oligonucleotides can be generated using chemical, physical or enzymatic techniques.

In another embodiment, the invention features a method of generating donor fragments for use in the RACHITT™ method, where the method take advantage of the activity of structure-specific endonucleases, *e.g.*, branch resolving enzymes, *e.g.*, T4 endonuclease VII or T7 endonuclase I. For example, the donor fragments for the RACHITT™ method can be generated via structure-specific endonucleases, *e.g.*, branch resolving enzymes, *e.g.*, T4 endonuclease VII or T7 endonuclase I. The action of such enzymes can be advantageously utilized by obtaining double-stranded nucleic acid to be used as the donor, incorporating uracil residues into a second double-stranded nucleic acid which is not intended to be the source of the donor fragments, treating both nucleic acids with lambda exonuclease to degrade the strand in each that contains a 5' phosphate, annealing the remaining strand from each to form a heteroduplex, treating the heteroduplex with an enzyme that recognizes mismatches (*e.g.*, a branch resolving enzyme such as T4 endonuclease VII or T7 endonuclase I) to yield homoduplexes of varying lengths, and treating the homoduplexes with an enzyme, *e.g.*, uracil DNA glycosylase, to degrade the strand containing the incorporated uracils, yielding single-stranded donor fragments. This is illustrated in Figure 15, which shows the double-stranded donor nucleic acid 430, and the second double-stranded nucleic acid 440, with uracils incorporated into it. Single-stranded polynucleotides are generated 450 and 460, respectively, which are annealed to form a heteroduplex containing mismatches 470. The heteroduplex is treated with an enzyme that recognizes mismatches to yield homoduplexes of varying lengths 480.

The size of the donor fragments generated with this method can be manipulated in a number of ways. The donor fragments can be size fractionated by any of a number of methods well-known in the art (*e.g.*, agarose gel electrophoresis). Duration of treatment with the branch resolving enzyme can also be used to alter the

size of the donor fragments produced. If known, the amount of mismatch between the two nucleic acid sources can also be determinative of the size of the homoduplexes, and therefore donor fragments, produced.

Because the amount of mismatch can determine the size of the donor fragments, the method described above selects for sequence diversity. The two source nucleic acids can be chosen on this basis, *e.g.*, the source genetic material can be two species related by varying degrees. Alternatively, a single source can be used, *e.g.*, a mixed culture from a soil sample, instead of two different nucleic acid sources.

By carefully selecting the source nucleic acids, this method can also be utilized to create donor fragments from specific areas of the genome, or from particular domains of a gene. For example, choosing source nucleic acids that are known to differ in sequence in a binding domain can produce donor fragments for the RACHITT™ method that are specific for that domain.

Moreover, any nucleic acid with similarity to the template can be used to generate the oligonucleotides of the present invention. As defined herein, sufficient similarity means that the sequence of the oligonucleotide need not reflect the exact sequence of the template. Conditions are chosen to allow such sequences (and those having low similarity or similar sequences interrupted with dissimilar sequences) to hybridize the template, such that chimeric polynucleotides are formed. For example, non-complementary bases or insertions or deletions can be interspersed in sequences.

As used herein, a variant of a polynucleotide or polypeptide refers to a molecule that is substantially similar to either the entire polynucleotide or polypeptide molecule, or a fragment thereof. For example, when the molecule is a polypeptide, variant refers to an amino acid sequence that is altered by one or more amino acids, wherein either a biological function, structure or antigenicity of said sequence or combination thereof is maintained in the variant. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, *e.g.*, replacement of leucine with isoleucine. Or a variant may have "nonconservative" changes, *e.g.*, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Similarly, when the molecule is a polynucleotide, variant refers to a sequence that is

altered by one or more nucleotides. The variant may have silent variations, wherein the change does not alter the amino acid encoded by the triplet comprising said variation or the variation is not silent, that is, alterations in encoded amino acids are generated.

As used herein, the term "altered version" refers to a polynucleotide sequence or a polypeptide sequence, wherein the nucleotide sequence has one or more differences in sequence from a native or wildtype version of said sequence. The term "native" or "wildtype" as used herein includes all alleles of the naturally occurring sequence.

As used herein, "homologs" refers to two or more sequences that share significant similarity but are not identical. Two DNA sequences are "substantially similar" when at least about 50%, preferably at least about 75%, and most preferably at least about 90 or 95% of the nucleotides match over the defined length of the DNA sequences, *e.g.*, when they share at least about 50%, preferably about 75%, and most preferably at least about 90 or 95% sequence identity over the defined length of the sequence. Sequences that are substantially similar can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, *e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., New York; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

The comparison of sequences and determination of substantial similarity, *e.g.*, percent identity, between two sequences can be accomplished using a mathematical algorithm (*Computational Molecular Biology*, Lesk, A.M.,ed., Oxford University Press, New York, 1988; *Biocomputing Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereaux, J., eds., M. Stockton

Press, New York, 1991). In a preferred embodiment, the "percent identity" between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*CABIOS, 4*:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. For purposes of the present invention, amino acid sequences having, for example, greater than 90 percent identity are considered substantially similar, *e.g.*, those with 90, 91, 92, 93, 94, 95, 96, 97, 98 and 99 percent identity.

Preferred methods for fragmenting parent polynucleotides in order to generate populations of oligonucleotides are those methods which produce fragments without particular sequence patterns. In one embodiment, chemical techniques of fragmenting polynucleotides include techniques that use conditions such as pH extremes, hydroxy radical formation, chemical radical formation and chemical catalysis. Methods of fragmenting oligonucleotides or polynucleotides using these chemical techniques are well known in the art. The methods of fragmenting by chemical techniques can be used to generate defined or random ends. For example, nucleic acid can be hydrolyzed after defined bases (such as after only guanosine). Nucleic acid can also be hydrolyzed, generating undefined termini. Such methods are well known in the art.

Methods of preparation can be selected to provide particular types of oligonucleotides, for example, in one embodiment, exposure to extreme pH, *e.g.* acid

pH or basic pH, can be used to generate fragments with random termini. In another embodiment hydroxyl radicals, *e.g.*, supplied using Fenton or Udenfriend reagent, react with the deoxyribose in DNA, resulting in cleavage of the DNA strand. In this embodiment, the upstream DNA product strand is left with a 3' phosphate, and the downstream DNA strand possesses a 5' phosphate. This method results in near uniform cleavage at any base within a target, and the frequency of cleavage can be regulated.

Physical techniques of fragmenting oligonucleotides include heat, freezing, ionizing radiation and shearing. Alternatively, enzymatic techniques of generating a random population of oligonucleotides or include use of any suitable enzyme such for example, a nucleic acid polymerizing enzyme or a nuclease. In one embodiment, a polymerase is used to synthesize oligonucleotides of variable length. Methods of such syntheses are well known in the art. For example, where fragments are generated by template dependent synthesis, conditions of synthesis can be chosen such that said polymerase falls off the polynucleotide from which said oligonucleotides serve as primers, or such that the polymerase otherwise terminates synthesis at random points along the polynucleotide. Oligonucleotides can be generated with random alterations. Oligonucleotides with random termini can be generated by using polymerases that are known to have exonuclease activity, under conditions permitting exonuclease activity for example, T4 DNA polymerase, Pol I or Pol III, among others.

Oligonucleotides can be produced by removing bases or generating adducts in a oligonucleotide using techniques well known in the art. For example, bases in oligonucleotides can be removed or adducted by many well known chemical methods to result in either abasic sites or modified bases. Incorporation of the uracil in a DNA polynucleotide, for example, can lead to the formation of abasic sites when treated enzymatically. The percent of uracils contained in a DNA polynucleotide can be controlled, and, thus, the number and average distance between uracil residues which eventually result in abasic sites can be regulated. These sites can be produced, for example, between 15 and 5000 bases apart (Kunkel *et al.*, *Meth. Enzymol.* *154*:367-382 (1987)). Strand cleavage of the phosphodiester bond at those modified

sites can then be effected using chemicals such as piperidine, or enzymes such as abasic endonucleases or lyases.

Oligonucleotides having defined ends can be generated using endonucleases having sequence recognition sites. Such enzymes are known in the art and referred to herein as "restriction endonucleases" and many are commercially available. A random population of fragments can be generated by performing a limited or incomplete digestion of the nucleic acid source of oligonucleotides. Oligonucleotides having random termini can be generated by using non specific endonucleases such as mung bean nuclease, S1 nuclease or DNase I. Alternatively, oligonucleotides having undefined ends can be generated, exonucleases such as Exo III can be used to non-specifically trim oligonucleotide sequences.

In a particular embodiment, a randomly fragmented population of oligonucleotides is generated using endonucleases which indiscriminately cleave single-stranded DNA, such as S1 nuclease or mung bean nuclease. In this embodiment, double-stranded parent polynucleotides are denatured and allowed to reanneal. If parents of related but not identical sequence are used, then, after annealing, double-stranded molecules will be formed with single-stranded mismatched "bubble" regions that form due to slight differences in sequence. These bubble sequences are susceptible to cleavage by single-stranded exonucleases.

An oligonucleotide of interest or a fragment thereof can be amplified *in vitro*. Such *in vitro* amplification includes enzymatic methods such as PCR and chemical synthesis. In a particular embodiment, short oligonucleotides (preferably from about 5 to 100 nucleotides in length) are synthesized *in vitro* to include specific polymorphisms contained on various parent polynucleotides. Alternatively, an oligonucleotide of interest or a fragment thereof can be amplified *in vivo* by, for example, plasmid or chromosomal replication in an appropriate host cell.

In one embodiment, the present invention is directed to a method of generating chimeric polynucleotides. As described herein, a random population of oligonucleotides can be generated, for example, by random fragmentation of polynucleotides of interest. In this embodiment, the method involves randomly assorting oligonucleotides on a single-stranded polynucleotide template based on

their ability to hybridize the template. The term "randomly assorting," as used herein, is intended to include those processes in which two or more oligonucleotides can hybridize to the template in a manner such that the possibility exists, that even though only one oligonucleotide actually hybridizes to a particular template sequence, that any one of two or more oligonucleotides could have hybridized to that particular sequence. It is clear to one of skill in the art that the method of the present invention can be carried out under a range of hybridization stringencies, from low stringency to high stringency, based on the amount of similarity or differences between the oligonucleotides and the template. In one embodiment of the present invention, the oligonucleotides are hybridized or annealed to the template (or templates) under conditions of low stringency.

Conditions of hybridization can be selected such that two or more oligonucleotides or portions thereof can compete for hybridization to a given region of a template and thus compete to be included in a resulting chimeric polynucleotide. A population of chimeric polynucleotides can be generated, representing a library of chimeras, which can be selected or screened to isolate chimeric polynucleotides with desired characteristics.

Moreover, the random assortment of oligonucleotides on a template to generate a chimeric polynucleotide can be facilitated by the population of oligonucleotides. A random population of oligonucleotides can be provided that hybridizes to two or more distinct regions of a template. A random population of oligonucleotides can be provided in which hybridization of two or more oligonucleotides terminates at different nucleotides of the template. Additionally, a random population of oligonucleotides can be provided that hybridizes to two or more distinct regions of a template and independently with respect to two or more regions, the hybridization of two or more oligonucleotides terminates at different nucleotides of the template.

A general description of stringency for hybridization conditions is provided by Ausubel, F.M., *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley-Interscience 1987, & Supp. 49, 2000, the teachings of which are incorporated herein by reference. Factors such as nucleotide length, base

composition, percent mismatch between the hybridizing sequences, temperature of reactions and ionic strength influence the stability of nucleic acid hybrids. Thus, stringency conditions sufficient to allow hybridization of oligonucleotides to the template, can be varied significantly and still allow oligonucleotides to hybridize and then generate at least one chimeric polynucleotide. Conditions sufficient to allow alignment between sequences of low similarity or sequences of high similarity interrupted with sequences of low or no sequence similarity can be low stringency conditions. Such conditions can readily be determined by one of ordinary skill in the art.

Specific consideration to hybridization and wash conditions is required in cases where relatively small oligonucleotides are being annealed, or in cases where there is a substantial amount of mismatch between the two strands. The energetics favoring hybridization indicate that longer stretches of homology are more favorable. Thus, when either short sequences are involved or there is limited potential for standard Watson-Crick base-pairing, hybridization conditions can be adjusted to a lower stringency to allow for hybridization. Typically, adjusting hybridization and wash conditions is done by, for example, adjusting the ionic strength of the reaction mixture or adjusting the temperature at which the hybridization is performed. In addition, certain purified proteins, such as the *E. coli* RecA protein, aid in homologous base pairing and can be included to facilitate hybridization of polynucleotide strands.

Conditions for stringency are as described in WO 98/40404, the teachings of which are incorporated herein by reference. In particular, examples of highly stringent, stringent, reduced and least stringent conditions are provided in WO 98/40404 in the Table on page 36. Examples of stringency conditions for solutions during and after hybridization are shown in the table below which is from WO98/40404 to Jacobs *et al.*, highly stringent conditions are those that are at least as stringent as, for example, conditions A-F; stringent conditions are at least as stringent as, for example, conditions G-L; and reduced stringency conditions are at least as stringent as, for example, conditions M-R.

Stringency Condition	Polynucleotide Hybrid	Hybrid Length (bp)[‡]	Hybridization Temperature and Buffer[†]	Wash Temperature and Buffer[†]
A	DNA:DNA	≥ 50	65°C; 1xSSC -or- 42°C; 1xSSC, 50% formamide	65°C; 0.3xSSC
B	DNA:DNA	<50	T_B*; 1xSSC	T_B*; 1xSSC
C	DNA:RNA	≥ 50	67°C; 1xSSC -or- 45°C; 1xSSC, 50% formamide	67°C; 0.3xSSC
D	DNA:RNA	<50	T_D*; 1xSSC	T_D*; 1xSSC
E	RNA:RNA	≥ 50	70°C; 1xSSC -or- 50°C; 1xSSC, 50% formamide	70°C; 0.3xSSC
F	RNA:RNA	<50	T_F*; 1xSSC	T_F*; 1xSSC
G	DNA:DNA	≥ 50	65°C; 4xSSC -or- 42°C; 4xSSC, 50% formamide	65°C; 1xSSC
H	DNA:DNA	<50	T_H*; 4xSSC	T_H*; 4xSSC
I	DNA:RNA	≥ 50	67°C; 4xSSC -or- 45°C; 4xSSC, 50% formamide	67°C; 1xSSC
J	DNA:RNA	<50	T_J*; 4xSSC	T_J*; 4xSSC
K	RNA:RNA	≥ 50	70°C; 4xSSC -or- 50°C; 4xSSC, 50% formamide	67°C; 1xSSC
L	RNA:RNA	<50	T_L*; 2xSSC	T_L*; 2xSSC
M	DNA:DNA	≥ 50	50°C; 4xSSC -or- 40°C; 6xSSC, 50% formamide	50°C; 2xSSC
N	DNA:DNA	<50	T_N*; 6xSSC	T_N*; 6xSSC
O	DNA:RNA	≥ 50	55°C; 4xSSC -or- 42°C; 6xSSC, 50% formamide	55°C; 2xSSC
P	DNA:RNA	<50	T_P*; 6xSSC	T_P*; 6xSSC
Q	RNA:RNA	≥ 50	60°C; 4xSSC -or- 45°C; 6xSSC, 50% formamide	60°C; 2xSSC
R	RNA:RNA	<50	T_R*; 4xSSC	T_R*; 4xSSC

[‡]: The hybrid length is that anticipated for the hybridized region(s) of the hybridizing polynucleotides. When hybridizing a polynucleotide to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing polynucleotide. When polynucleotides of known sequence are hybridized, the hybrid length can be determined by aligning the sequences of the polynucleotides and identifying the region or regions of optimal sequence complementarity.

†: SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete.

*T_B - T_R: The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}[Na^+]$) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and $[Na^+]$ is the concentration of sodium ions in the hybridization buffer ($[Na^+]$ for 1xSSC = 0.165 M).

It is clear to one skilled in the art that this hybridization step can be optimized using any suitable method of optimization that is established in the art of hybridization and or PCR. These include, but are not limited to, techniques that increase the efficiency of annealing or hybridization from complex mixtures of polynucleotides (*e.g.*, PERT; Nucleic Acids Research, 1995, vol 23, 2339-2340) or hybridization in different formats (*e.g.*, using immobilized template or in microtiter well formats; Analytical Biochemistry, 1995, vol 227, 201-209).

Oligonucleotides can hybridize to the template such that the entire length of the oligonucleotide is hybridized to the template. Alternatively, oligonucleotides can hybridize to the template such that portions remain unhybridized, *e.g.*, at least one terminus remains unhybridized, *e.g.*, one terminus forms a "flap." The term "flaps" is used herein to describe the unhybridized portion of an oligonucleotide otherwise hybridized to a template. Internal sequences can also remain unhybridized, thus forming "loops" (loops are observed, for example, during denaturation/renaturation experiments with cDNA and genomic genes in which genomic introns loop out since there is no corresponding cDNA sequence to which to hybridize). In embodiments in which one or more oligonucleotides form flaps, the method typically includes the further step of removing or cleaving flaps from the oligonucleotide prior to the ligation step. In embodiments where one or more oligonucleotides form loops, the method typically includes the further step of removing or trimming loops from the template or oligonucleotide prior to the ligation step. The "trimming" of flaps or loops, used herein to refer to a process of removing just the flaps or loops and leaving the hybridized portion of the

oligonucleotide intact, can be incorporated into the method of the present invention. In embodiments where one or more oligonucleotides form both flaps and loops, these structures can both be trimmed prior to the ligation step. Flaps and loops can be trimmed enzymatically, *e.g.*, utilizing polymerases with single-stranded exonuclease activity or other single-stranded endonucleases or exonucleases (*e.g.*, mung bean nuclease, S1 nuclease), or chemically. In another embodiment, the step of trimming flaps and/or loops can be performed prior to or concurrently with the additional steps of polymerization and ligation.

For example, flaps can be removed using RNA or DNA digesting enzymes, including nucleic acid polymerases having proof reading ability. Flaps can also be removed using enzymes including, but not limited to, RNase HI, Exo III, Exo VII or the flap endonuclease activity of *Taq* DNA polymerase. Methods of using these enzymes are well known in the art. Enzymes can be selected on their abilities to perform at various temperatures, for example, they can be cryophilic in that they are active at low temperatures (less than about 30° C), mesophilic in that they are functional at moderate temperatures (between about 30° C and 60° C), or thermophilic in that they are active at high temperatures (greater than about 60° C).

In one embodiment including a step of removing flaps of the hybridized oligonucleotides, unhybridized termini of the template can be protected from degradation by including within the population of oligonucleotides of interest nucleic acid sequence specific for the termini of the template. These fragments, termed "clamps," or equivalently, "anchors," can be from about 6 to about 500 or more nucleotides in length. In a particular embodiment, the clamps can be from about 12 to about 50 nucleotides in length. These clamps can be designed to have low stringency sufficient to protect the termini from degradation but such that oligonucleotides of interest that bind the termini can displace the clamps. In another embodiment, the termini of the template can be protected by having one or more thioester or phosphorothioate bonds instead of a phosphodiester bond, thus rendering the termini refractory to trimming enzymes. In another embodiment, the template comprises, for example, PNA, which is unaffected by nucleic acid polymerizing or degrading enzymes.

Oligonucleotides can hybridize to the template such that no region of the template remains unhybridized, *e.g.*, there is no intervening single-stranded template sequence between the hybridized oligonucleotides, *e.g*, there are no gaps between the hybridized oligonucleotides. Alternatively, oligonucleotides can hybridize to the template or templates such that regions of the template remain unhybridized, *e.g.*, there is intervening single-stranded template sequence between the hybridized oligonucleotides *e.g.*, there are gaps between the hybridized oligonucleotides. The term "adjacently hybridized" is used herein to describe the relative positions of two oligonucleotides hybridized to the same template at positions such that only single-stranded template sequence is contained between the two oligonucleotides. The term "immediately adjacently hybridized" is used herein to describe adjacently hybridized oligonucleotides which abut each other, *e.g.*, no intervening single-stranded template is contained between the two oligonucleotides.

In embodiments in which templates contain gaps after hybridization, the method includes the further step of filling in the gaps between hybridized oligonucleotides, *e.g.*, rendering the unhybridized regions of the template double-stranded, such that the step of ligating the hybridized oligonucleotides results in a chimeric polynucleotide about target length.

Trimming of flaps and gap filling can be performed in any order, including both sequentially and concurrently. Moreover, ligation can also be performed concurrently with trimming of flaps and gap filling. The three process steps in combination are referred to as the TPL step. Thus, the TPL step can include trimming flaps, polymerization to fill in gaps between adjacently hybridized oligonucleotides, and ligation to chemically join immediately adjacently hybridized oligonucleotides. Typically, the TPL step of the invention follows the contacting and hybridization of the population of oligonucleotides to the single-stranded template.

The present invention is also related, at least in part, to the discovery of modifications of single-stranded templates used to assemble oligonucleotides prior to recombination. In one embodiment, the present invention optimizes the TPL step. Such modifications to the single-stranded template lead to an increase in the

degree of chimeragenesis in the products by limiting polymerization during the TPL step to filling in only short gaps between adjacently hybridized oligonucleotides. Since filling in gaps between oligonucleotides results in the synthesis of sequences complementary to the template, the overall degree of chimeragenesis of the eventual product is typically lowered when long gaps are filled in. Moreover, this improvement also increases the overall proportion of chimeric products formed by limiting the formation of non-chimeric products. Indeed, in the extreme case, polymerization can lead to the formation of a non-chimeragenic product, *e.g.*, the synthesis of a strand entirely complementary to the template. Such a non-chimeragenic product can be produced by a single oligonucleotide hybridizing to the template and then being extended for the length of the template during the polymerization step. Thus, limiting the polymerization step to short gaps can increase the overall degree of chimeragenesis of the product polynucleotide. Modifications to the single-stranded template include both physical and chemical modifications.

In one embodiment, the method of the present invention limits the polymerization step to filling in only short gaps (*e.g.*, gaps that span less than about 5%, 10%,15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, and 95% of the template sequence) and, thus, further minimizes the extensive polymerization that leads to non-chimeragenic product polynucleotides or polynucleotide products with a low degree of chimeragenesis.

In one embodiment of the present invention, the further step of filling in gaps between hybridized oligonucleotides can be conducted before removal of the unhybridized sequences and thereby protect the 3' terminus of the template. After hybridization of oligonucleotides to templates, gaps between adjacently hybridized oligonucleotides can be filled using a suitable nucleic acid polymerizing enzyme. Suitable polymerases are readily available. Gaps can be filled in using prokaryotic, eukaryotic or viral polymerases. The polymerases can be selected on their abilities to perform at various temperatures, for example, they can be cryophilic in that they are active at low temperatures (less than about 30° C), mesophilic in that they are functional at moderate temperatures (between about 30° C and 60° C), or

thermophilic in that they are active at high temperatures (greater than about 60° C).. The polymerases can optionally have proof reading ability. Suitable polymerizing enzymes include but are not limited to T4 DNA polymerase, *Taq* DNA polymerase, Pfu DNA polymerase, Pol I, Klenow and Klenow 3'-5'$^{\text{exominus}}$ (New England BioLabs, Beverly, Ma.).

Control of enzymatic polymerization can be achieved, for example, by affecting the polymerase, *e.g.*, by altering reaction conditions using a polymerase with altered processivity, or by affecting the template which is used by the polymerase during polymerization.

The single-stranded template can be modified such that polymerization during the TPL step of the invention is affected. The modification can be either chemical or physical. For example, the template can include the nucleotide, uracil, which can be glycosylated by a uracil glycosylase, "UNG," thus forming an abasic residue within the polynucleotide template. Known polymerases do not recognize these abasic residues and stall upon encountering such residues during synthesis. By controlling the number of uracils in the template, the extent of polymerization can be modulated by affecting the average distance a polymerase can cover before encountering a modified residue. Other chemical modifications to template polynucleotides which lead to stalled polymerases are known to those of ordinary skill in the art.

Additionally, the template can be modified by annealing pre-selected oligonucleotides to the template. For example, an oligonucleotide complementary to a template terminus can be hybridized to the template, thus preventing oligonucleotides in the arbitrarily fragmented population of oligonucleotides generated from parent polynucleotides from hybridizing to the terminus of the template. In one embodiment of the present invention, the pre-selected oligonucleotide cannot be extended by polymerases, *e.g.*, it can contain a 3' dideoxy or 3' phosphate containing nucleotide. Under these conditions, the polymerase utilized for filling in gaps is prevented from synthesizing a complementary and non-chimeragenic copy of the template.

After gaps are filled in, the pre-selected oligonucleotide can be selectively removed and replaced by a second pre-selected oligonucleotide which can be extended by polymerases, or a second pre-selected oligonucleotide which can be extended by polymerases can be immediately adjacently hybridized to the first pre-selected oligonucleotide. For example, the first pre-selected oligonucleotide can be synthesized to contain the nucleotide, uracil, instead of thymidine. If the second, extendable oligonucleotide does not contain uracil but instead contains thymidine, conditions can be established such that after UNG treatment the second oligonucleotide will be substantially favored over the first oligonucleotide for annealing to the template. Alternatively, a cleavage site for a restriction endonuclease can be contained at the 3' end of the template. Thus, after the TPL step, the template/chimera hybrid can be treated with the endonuclease, thus leaving a 3' overhang and a short portion of the first pre-selected oligonucleotide. This short portion can be removed while other oligonucleotides remain hybridized by adjusting the stringency of the hybridization. At this point, the second pre-selected oligonucleotide is specifically annealed to the 3' end of the template. In either case, a subsequent round of polymerization to extend the second pre-selected oligonucleotide and ligation generates the full length chimeric polynucleotide.

Gaps between adjacently hybridized oligonucleotides can be of any length. For example, gaps between adjacently hybridized oligonucleotides can be from about 1,000 to about 100,000 template nucleotides. The gaps between adjacently hybridized oligonucleotides can be from about 500 to about 10,000 template nucleotides; less than 1,000 template nucleotides; less than 250 template nucleotides; less than 50 template nucleotides; or even less than 25 template nucleotides.

In another embodiment, gaps are filled in *in vivo*, wherein complexes comprising oligonucleotides hybridized to a template are inserted or transformed into a suitable host cell. For the purposes of the present invention, the gaps, whether filled in *in vitro* or *in vivo*, can be filled with or without the introduction of "errors" in comparison to the template.

In another embodiment, the invention is directed to a method of generating a chimeric polynucleotide having at least one characteristic altered in comparison to a reference polynucleotide including the steps of generating a population of oligonucleotides, comprising randomly fragmenting at least one polynucleotide; contacting at least one single-stranded polynucleotide template with the population of oligonucleotides under conditions wherein at least two oligonucleotides hybridize to the template; removing flaps of hybridized oligonucleotides; filling in gaps between hybridized oligonucleotides; ligating adjacently hybridized oligonucleotides to form at least one chimeric polynucleotide, hybridized to a template; selectively amplifying the chimeric polynucleotides and selecting or screening at least one chimeric polynucleotide having the specified characteristic altered in comparison to the reference polynucleotide.

Typically in the method of the present invention, a chimeric polynucleotide is produced in a single iteration of the steps of the method. That is, after oligonucleotides hybridize to the single-stranded template, and any flaps or loops are trimmed and any gaps remaining are filled in, and the hybridized oligonucleotides are ligated, a chimeric polynucleotide is formed which is usually about the length of the template. Individual process steps are generally not repeated, nor are the chimeric polynucleotide products formed in stepwise extensions. The necessary steps of the process can, however, be repeated after the chimeric polynucleotide is formed, for example, the chimeric polynucleotide can be utilized as a library of oligonucleotide fragments or as a template in a subsequent reiteration of the process. Thus, the methods of the invention are generally considered to constitute a non-reiterative process.

Moreover, as described previously, the chimeric polynucleotide product formed is generally about the length of the template on which it is assembled. Although a chimeric strand may not form on any given template, about one polynucleotide product is produced per single-stranded template upon which a chimeric strand is synthesized, and the number of polynucleotide products produced and the number of such templates are in a ratio of about 1.

In one embodiment of the present invention, the chimeric polynucleotides are selectively amplified with respect to the template.

In the method of the present invention, at least one chimeric polynucleotide is generated, wherein one or more characteristics of the chimeric molecule is altered in comparison to at least one reference molecule. For example, parent polynucleotides used to generate the population of oligonucleotides contain sequence differences which account for differences in characteristics associated with the parent polynucleotides, *e.g.*, different parent polynucleotides can have slight differences in the nucleotide sequences of their coding regions wherein such differences result in the expression of an altered protein with altered functions. These sequence differences are referred to as "polymorphisms", and, in cases where the parent polynucleotides contain reading frames, such different parent polynucleotides can represent different "alleles". When such parent polynucleotides are fragmented, oligonucleotides containing specific polymorphisms derived from specific parents are generated. The random set of oligonucleotides that hybridizes to any given template determines the precise set of polymorphisms contained to the chimeric product. A comparison of a chimeric product to any one of the parents or to a product of a polynucleotide functionally related to the parents allows for the evaluation of the chimeric polynucleotide product with respect to a given function. The molecule used for such a comparison is referred to herein as a "reference" polynucleotide.

It is understood that the methods of the present invention can comprise any combination of the embodiments described.

The nucleic acid which forms the parent polynucleotides, or that which is utilized as a template or in a population of oligonucleotides can be isolated from any suitable source of polynucleotides or it can be synthetically manufactured. For example, it can be isolated from an organism, such as a eubacterial, archeal, eukaryotic or viral organism. These organisms can be amplified, enriched or isolated and grown in culture, or can be used directly from environmental sources.

Environmental sources include soil samples, water samples from fresh water sources or salt water sources, polluted sites, waste treatment sites and sources

comprising extreme condition sources such as permafrost sources, high altitude sources, high pressure sources and geothermal sources such as volcanic sources, hot springs and hydrothermal vent sources. Sources of nucleic acid also include tissue or bodily fluid samples from an organism, wherein nucleic acid can be isolated from said sample. The nucleic acid of a tissue or bodily fluid sample can comprise nucleic acid of said organism or of flora, such as fungal, bacterial, viral or parasitic organisms present in said sample. The sample can further be fresh, fossile or archival. One source of nucleic acid for such oligonucleotides is polynucleotide encoding a desulfurization enzyme.

These isolated nucleic acids can comprise a variety of forms of nucleic acids including both genomic DNA and cDNA. For example, in the case of a cDNA library, RNA, or more particularly mRNA can be isolated from a nucleic acid containing sample and converted into DNA, *e.g.* cDNA, according to standard methods. The library can be prepared from a sample of interest that expresses a desired phenotype. Alternatively, the library can be produced from a sample that is not otherwise preselected for phenotype.

In one embodiment, the library is subjected to subtractive hybridization against a suitable sample of nucleic acid, using techniques well known in the art. A suitable sample of nucleic acid includes nucleic acid from a reference strain of bacteria, for example, one that does not express the phenotype of interest. Sequences that are common between the library and the sample are removed from the library. In this way, sequences present in multiple copies and sequences that are common between the two populations are removed, effectively enriching for low abundance sequences. Particular methods for producing libraries of oligonucleotides with minimal complementarity are described in the application entitled "Generating Single-Stranded Oligonucleotide Libraries With Minimal Complementarity And Uses Therefore" by Joseph J. Arensdorf and Wayne M. Coco with Attorney Docket No. 1405.2016-001 filed of even date herewith, the entire teachings of which are hereby incorporated by reference.

In one embodiment, a region of the template is hybridized to complementary sequences by providing oligonucleotides complimentary to any oligonucleotide or

oligonucleotides that are complementary to that region of the template. In this manner, the sequence complementary to a region of the template can be specifically retained in the chimeric molecule. Conversely, defined oligonucleotides can be added in greater quantities to the random population of oligonucleotides in order to preferentially hybridize the defined oligonucleotides to the template. In one embodiment these oligonucleotides hybridize at particular regions or positions in order to introduce desired mutations or in order to protect sequences on the template from changes that might be introduced by the random oligonucleotides.

The present invention allows oligonucleotides of interest to be incorporated into a larger molecule to form one or more chimeric polynucleotides, hybridized to a given template. In one embodiment, polynucleotides that are not otherwise easily manipulated (due for example, to their large size), can be separately manipulated as oligonucleotides and rejoined by hybridizing the oligonucleotides to a transient template. For example, random mutagenesis using PCR is most effective on smaller DNA fragments, such as 1 kilobase or less in length. In one embodiment, a large polynucleotide can be cleaved into fragments of about 1 kilobase and randomly mutagenized using PCR, and then made single-stranded, fragments can be hybridized to a transient template and ligated as described herein. The transient template can be derived from the original polynucleotide, or can be an altered form thereof as described herein, for example, the template itself can be mutagenized or can have added or deleted regions or domains as compared to the starting polynucleotide.

In a particular embodiment, a set of specific oligonucleotides incorporating known allelic variations can be synthesized and used as the population of oligonucleotides in the RACHITT™ method. In this embodiment, polymorphisms which are located in positions close to each other on the nucleotide sequence of the parent polynucleotide can be sampled separately based on available sequence and allele information. The particular combination of polymorphisms contained in the chimeric polynucleotide resulting from this method is determined by the specific set of oligonucleotides which hybridizes to the template during the RACHITT™ method. As such, the set of polymorphisms contained in each chimeric

polynucleotide is a random set which can vary from chimera to chimera, thus allowing production of a library containing many different chimera.

Certain steps of the present invention can be conducted using *in vitro* techniques, *in vivo* techniques or a combination thereof. In one embodiment of the present invention, *in vivo* techniques can include methods for trimming of unhybridized termini, gap filling, ligation, or combinations thereof, wherein the oligonucleotides hybridized to the template are transformed into a suitable host cell. Suitable host cells include bacteria and eukaryotic cells, such as yeast, mammalian cell culture, insect cell culture, plant or in multicellular organisms. Suitable bacterial host cells include, for example, bacteria that express desulfurization biocatalysts, and non-desulfurization biocatalyst expressing bacteria, such as *E. coli*, *Bacillus* and non-desulfurizing pseudomonads.

Suitable host cells for *in vivo* trimming of unhybridized termini, gap filling or ligation, or for cloning, selecting, screening, expressing the chimeras of the present invention include bacteria that express native or altered desulfurization biocatalyst such as *Sphingomonas* microorganisms such as *Sphingomonas* strain AD 109, *Rhodococcus* microorganisms such as *Rhodococcus* strain IGTS8, *Corynebacterium* such as *Corynebacterium* strain SY1, *Rhodococcus erythropolis D-1* and ATCC #55309 and #55310, *Arthrobacter* described by Lee *et al.*, *Appl. Environ. Microbiol.* 61:4362-4366 (1995) *Agrobacterium* described by Constanti *et al.*, *Enzyme Microb. Tech.* 19:214-219 (1996) and *Mycobacterium*, such as *Mycobacterium* strain G3 described in *Biocatalysis and Biotransformation*, *15*:17-27 (1996). Suitable host bacteria also include for example, *E. coli*, *e.g.*, K12 and its derivatives and *Bacillius*. Suitable host selection and/or screening can also be performed *in vitro*.

Furthermore, the steps of generating a random population of oligonucleotides, gap filling, ligation, or combinations thereof including chemical or enzymatic nucleic acid hydrolysis, gap filling, ligation or combinations thereof can be performed using *in vitro* techniques.

In the method of the present invention, hybridized oligonucleotides are ligated. The hybridized oligonucleotides to be ligated are adjacently hybridized on

the template. The hybridized oligonucleotides are ligated using a suitable ligase enzymes. In one embodiment, oligonucleotides are ligated using one or more ligases. Suitable ligases include both thermostable and non thermostable ligases and include T4 DNA ligase, DNA ligase I or *Taq* ligase and *Tth* ligase. Hybridized oligonucleotides can also be ligated using chemical means.

Although variations are contemplated, typically the methods of the invention produce a single contiguous polynucleotide of about the same length as the template. Thus the chimeric polynucleotides can be described as being formed in about a 1:1 ratio with the transient templates of the invention. A plurality of chimeric polynucleotides can be formed on a plurality of transient templates.

The method of the present invention further comprises repeating the method using at least a chimeric polynucleotide or fragment thereof formed in an earlier process as the hybridization template or source of hybridizing oligonucleotides.

In still another embodiment of the present invention, the chimeric polynucleotides are selectively amplified with respect to the template. Typically, the transient template is not directly incorporated into the resulting chimeric molecules. Specific amplification of the chimeric polynucleotide product can be followed by one or more rounds of selection or screening and subsequent amplification. The chimeric polynucleotide can be selected or screened based on alterations of specific properties, *e.g.*, nucleotide structure, nucleotide function, altered enzymatic activities of proteins encoded by the chimeric polynucleotide, altered structural functions of proteins encoded by the chimeric polynucleotide, altered regulatory functions of proteins encoded by the chimeric polynucleotide, etc., or a combination thereof. Subsequent selection and amplification of the chimeric polynucleotide allows for the *in vitro* or *in vivo* directed evolution of biological molecules such as nucleic acid or polypeptides. This method for directed evolution can aid in the improvement of such molecules for use, for example, in medical therapies, as reagents in molecular biology, and in industry.

A template can be synthesized such that after ligation to form the chimeric polynucleotides, the chimeric polynucleotide is selectively amplified over the template. In one embodiment, the template is synthesized to contain at least one

uracil. In this embodiment, the template is rendered non-amplifiable or non-replicatable by treating the template with Uracil DNA Glycosylase (UDG). In another embodiment, the template-chimera heteroduplex can be introduced into a suitable host cell, wherein the template is selectively destroyed. A suitable host can be a bacterial strain carrying the wild type *dut* gene. Methods of transfecting suitable hosts, such as bacteria, are well known in the art. Moreover, the chimeric polynucleotide and template can be inserted into a suitable vector to facilitate the process.

Alternatively, to facilitate selectively amplifying the chimeric polynucleotides over the templates, the template or templates can be synthesized such that they can be separated by differential mobility, density, or charge; or immobilization on a matrix or solid phase. In still another embodiment, the template is synthesized such that it contains an affinity moiety, such that it can be separated by selective binding or removal, such as affinity chromatography. In one embodiment, the affinity moiety comprises, for example, biotin or digoxygenin, which can be affinity separated along with the molecule to which it is attached, using avidin or anti-digoxygenin antibodies, respectively. In yet another embodiment, one or more of the oligonucleotides of interest or fragments thereof can be synthesized to contain an affinity moiety such that the resulting chimera contains the moiety. The chimeric polynucleotide can then be isolated as described above.

In another embodiment, the transient template used in the RACHITT™ method is formed from RNA. The initial chimeric product produced by this embodiment is a RNA/DNA hybrid in which the DNA strand is the chimeric strand. Thus, in this embodiment, the RNA strand can be selectively removed by ribonucleases such as RNase H, and the chimeric strand can then be selectively amplified. Alternatively, the chimeric DNA strand can be selectively amplified by PCR when PCR utilizes DNA-dependent DNA polymerases which do not recognize an RNA strand.

The transient template used in the RACHITT™ method can also contain sequences in addition to target sequences. Thus, the product of the RACHITT™

method can be a heteroduplex wherein the template strand is larger than the chimeric strand. In this embodiment, the template is selectively removed by methods known in the art (*e.g.*, electrophoresis, gel filtration, affinity chromatography, ion-exchange chromatography, subtractive hybridization), and the chimeric strand is then selectively amplified. In a particular embodiment, the transient template is a circular molecule and can contain sequences derived from phage genomes in addition to RACHITT™ target sequences. In another embodiment, the transient template contains sequences designed specifically to make the template larger than the resultant chimeric strand. Such a template can contain more than one copy of the target sequence, which can be separated by additional sequences.

In still another embodiment, the template can comprise PNA, which cannot be amplified using polymerases, allowing selective amplification of the chimeric polynucleotides.

In one embodiment, selection of chimeric polynucleotides with improved properties with respect to a particular trait is performed after cloning the library of chimeric products into a suitable vector. In one embodiment, restriction sites for cloning into suitable vectors can be contained at the 3' and 5' ends of the template. In this embodiment, the chimeric products can retain those restriction sites and, therefore, can be directly cloned into the same sites in a suitable vector.

One advantage for including a restriction site at the 5' end of the template is that only molecules made double-stranded during the TPL step of the RACHITT™ method can be cleaved at the restriction site. When the TPL step of the RACHITT method is performed without an extendible pre-selected 3' anchoring oligonucleotide, the restriction site at the 5' end of the template is made double-stranded only in the case where at least one oligonucleotide from the library of oligonucleotides hybridizes to the template. Treatment with the endonuclease that specifically cleaves DNA at the 5' site yields a 5' clonable end. Subsequent use of an extendible 3' oligonucleotide containing a different restriction endonuclease site followed by an additional polymerization, ligation and treatment with the endonuclease that specifically cleaves the 3' site, yields a molecule with a clonable

3' end. After these steps, only products that have incorporated at least one parent-derived oligonucleotide into the non-template strand will have both a 5' and a 3' clonable end. Undesirable "read-through" products resulting from polymerization of the entire template from a 3' anchoring oligonucleotide will not have a 5' clonable end. Thus, these undesired products are lost during a subsequent cloning step into a vector utilizing both the 5' and 3' restriction sites as cloning sites.

The RACHITT™ method provides a method for generating chimeric polynucleotides by means of *in vitro* directed evolution. Different sequences from related parent polynucleotides can be incorporated into the chimeric product. As such, the products formed represent a sample of the differences in the sequences of different parent polynucleotides. Typically, the RACHITT™ method results in a library of chimeric products, each chimera containing a different set of sample sequences derived from parent polynucleotides. A significant contaminant in this library can be a molecule which represents a clone of a parent sequence, *i.e.*, a molecule which contains sequence from only one parent or reference molecule, typically the template used in the RACHITT™ method. Reducing the occurrence of these parental clones improves the diversity of the library of chimeric products obtained as a result of the RACHITT™ method, and, thus, allows for a greater and more efficient sampling of combined parent sequences. A common pathway which leads to the appearance of parental clones in the final library of chimeric products involves a template which is extended during a polymerization step, but which does not contain parent-derived sequences annealed to it. An avoidance of such "read-through" products which are parental clones significantly improves the efficiency of the RACHITT™ method.

Thus, the RACHITT™ method can be modified to include, either before or during the hybridization step, pre-selected oligonucleotides which hybridize to either end of the template. These oligonucleotides can specifically optimize the RACHITT™ method depending on the precise sequence and other physical properties of the pre-selected oligonucleotides. Figure 16A illustrates particular instances of such oligonucleotides. A given gene of interest 510, *i.e.*, a "target sequence" for the RACHITT™ method, can be amplified by PCR using a set of

primers such as PFT10 520 and RT1 530. These primers contain sequences which can hybridize to the end of the target sequence, but also contain sequences on their 5' ends which are engineered and not part of the target sequence. The non-target sequences of PFT10 and RT1 are contained in FT8 540 and RT4 550, respectively. FT11P 560 is identical to PFT10 except that FT11P contains three nucleotides 570 which do not hybridize to FT8 (thus, annealing FT8 to a PCR fragment amplified with FT11P as a primer, results in FT8 annealing to the 5' end of the FT11P-containing molecule, but a 3' flap that is not hybridized, and that corresponds to the three nucleotide mismatch, is generated). FT11P also contains a 3' phosphate and a phosphorothioate linkage between the ultimate and penultimate nucleotides. Likewise, RT2 580 is identical to a sequence complementary to the 5' end of RT1, except for the analogous three nucleotide mismatch 590 internal to RT2. RT3 600, which contains uracil residues instead of thymine residues, is complementary to the 3' end of RT1 and, if hybridized to RT1-containing sequences with RT2, produces a double-stranded region covering the entirety of RT1 with a nick between RT3 and RT2 and a three nucleotide bubble created by the mismatch in RT2.

Where a target sequence is amplified by PFT10 and RT1 670, the target sequence is extended at each end by additional sequences as described above. Figure 16B presents two possible outcomes of the RACHITT™ method using such a PCR-generated template. One outcome 610 is the result of the RACHITT™ method in the case where no oligonucleotides from the fragmented population of oligonucleotides hybridizes to the template. The other 620 is the result of at least one oligonucleotide from the fragmented population of oligonucleotides hybridizing to the template.

The RACHITT™ method is altered such that FT11P 560 and RT2 580 are annealed to the ends of the template either prior to or during the addition of the library of oligonucleotides 630. During the TPL step of the RACHITT™ method, FT11P is not extended due to the 3' phosphate moiety. After TPL, RT3 is added under conditions such that it can hybridize to a single-stranded DNA sequence complementary to itself, and the 3' phosphate is then removed from FT11P 640.

Subsequent gap fill in results in a double-stranded molecule, one strand of which is the original template 650 and 660.

Where at least one oligonucleotide from the population of oligonucleotides anneals to the template, RT3 is prevented from annealing to the single-stranded template, and, thus, the resultant double stranded molecule does not contain uracil residues 660. However, in the case in which no oligonucleotides from the population of oligonucleotides anneals to the template, *i.e.*, resulting in a non-chimeric double-stranded copy of the template, the non-template strand contains uracil residues due to the fact that RT3 was permitted to anneal to the template when added 650. If the products are amplified by PCR using primers FT8 and RT4, only molecules which do not contain uracil residues can be amplified, *i.e.*, the non-chimeric double-stranded template copies cannot be amplified and the chimeric strands can be amplified.

The present invention is particularly useful for evolving industrially or medically useful molecules or biochemical pathways, wherein the chimeric polynucleotide is itself a useful molecule (*e.g.*, promoter, aptamer, catalyst, enhancer or other regulatory element) or wherein the chimeric polynucleotide encodes a useful gene product. The chimeric polynucleotides can be or encode, for example, molecules that are more active under desired conditions or have altered or enhanced specificity, mutagenicity or fidelity. For example, desired conditions include conditions to which the reference molecule, oligonucleotides or template, or polypeptide encoded therein are not typically exposed or extreme conditions. Such extreme conditions include high or low temperature, extreme high or low pH, extreme salt concentration or ionic strength or extreme solvent conditions such as organic solvent conditions.

Examples of industrially or medically useful polypeptides or polynucleotides are well known in the art. In one embodiment of the present invention, the template is selected or designed to encode a molecule that is to be evolved.

Medically useful molecules include bioactive molecules. As used herein the term bioactive agent includes peptides, proteins, polysaccharides and other sugars, lipids, nucleic acid sequences, such as genes, and antisense molecules. Nucleic acid

encoding enzymes that produce, modify or degrade polysaccharides, other sugars and lipids can be used as the template, oligonucleotides or reference polynucleotide. Specific examples of bioactive molecules include but are not limited to: insulin, erythropoietin, interferons, colony stimulating factors such as granulocyte colony stimulating factor, growth hormones such as human growth hormone, LHRH analogs, LHRH antagonists, tissue plasminogen activator, somatostatin analog, r Factor VIII, r Factor IX, calcitonin, dornase alpha, polysaccharides, AG337, bone inducing protein, bone morphogenic protein, brain derived growth factor, gastrin 17 immunogen, interleukins such as IL-2, PEF superoxide, permeability increasing protein-21, platelet derived growth factor, stem cell factor, thyrotropin alfa and somatomedin C.

One of skill in the art would readily be able to select or design a template to encode the molecule of interest to be evolved according to the method of the present invention. Methods for measuring activity of hormones, interleukins, growth factors and angiogenesis inhibitors and the like under desired conditions are well known in the art. One of ordinary skill in the art can readily determine the activity of the hormone, interleukin, growth factor or angiogenesis inhibitor encoded by the chimeric polynucleotide produced by the present invention and select those having the desired characteristics.

Examples of such medically useful molecules to be evolved according to the present invention also include enzymes that synthesize drugs, antibiotics, vitamins or cofactors. Other examples include vectors and genes for gene therapy. In addition, molecules that have desired therapeutic effect can be altered to alter, *e.g.*, increase or decrease, toxicity, antigenicity or other side effects.

Methods for determining activity under desired conditions include standard methods well known in the art. One of ordinary skill in the art can readily determine the activity of an enzyme encoded by a chimeric polynucleotide and select those having the desired characteristics. Enzymes include but are not limited to fermenting enzymes, proteases, lipases, oxidoreductases such as alcohol dehydrogenase, polymerases, hydrolases and luciferase.

Examples of industrially useful molecules include enzymes that synthesize polyketides, transform small molecules, hydrolyze substrates, replace steps in organic synthesis reactions or degrade pollutants such as aromatic hydrocarbons (*e.g.*, benzene, xylene, toluene and naphthalene), polychlorinated biphenyls and residual herbicides and pesticides. Catabolic pathways can be evolved using the present invention such that enzyme pathways are produced that degrade manmade pollutants that otherwise are not, or are only slowly catabolized by wild type organisms. Polynucleotides encoding said molecules or fragments thereof can be used in the present invention as either the template, the oligonucleotides, the reference molecule, or combinations thereof.

A particularly useful metabolic pathway for the directed evolution process of the present invention includes enzymatic pathways that catalyze one or more steps in the desulfurization of thiophenes, such as benzothiophenes or dibenzothiophenes, as described in United States Patent Applications, Serial Numbers 09/030,995 and 08/851,089 and United States Patent No. 5,952,208, the teachings of which are incorporated herein by reference in their entirety. Polynucleotides encoding the molecules or fragments thereof can be used in the present invention as either the template, the oligonucleotides, the reference molecule, or combinations thereof. In a more particular embodiment, said nucleic acid source includes SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, and homologs, variants, analogs and fragments thereof. The nucleic acid source further includes nucleic acids encoding polypeptide sequences SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, variants, analogs and fragments thereof. For example, $DBTO_2$ monooxygenase genes such as *dszA* genes can be evolved according to the method of the present invention to possess a higher rate of enzyme activity, altered substrate range (such as, substituted or unsubstituted dibenzothiophene, benzothiophene, thiophene, alkyl sulfides, aryl sulfides, cyclic sulfides or other sulfur or non-sulfur containing molecules) or altered K_m.

In one embodiment, the method of the present invention is performed on a dibenzothiophene monooxygenase gene, wherein the dibenzothiophene monooxygenase gene is used as a nucleic acid source of at least one template,

oligonucleotide, reference molecule or combination thereof, wherein a chimeric polynucleotide is isolated that encodes a polynucleotide that converts DBT to the compound DBTO, such that the second reaction present in native dibenzothiophene monooxygenase, converting DBTO to $DBTO_2$, does not occur. In another embodiment, a *dszC* gene is evolved by the method of the present invention such that it catalyzes the production of chiral sulfoxides in greater enantiomeric excess than that provided by the reference *dszC* gene. Methods of measuring DBTO production and activity of the desulfurization genes and chimeric genes of the present invention are found in United States Patent Application Serial Number 09/392,846 the teachings of which are incorporated herein by reference in their entirety.

The method of the present invention can be used to increase, for example, the rate of an enzyme activity and the extent of the activity, *e.g.*, the affinity of the enzyme for its substrate. For example, the first enzyme in the metabolism of sulfur heterocycles by *Rhodococcus,* dibenzothiophene-monooxygenase (DBT-MO), is the bottleneck for both the rate and extent of sulfur oxidation in the biodesulfurization (BDS) process. The method of the present invention was used to generate large, complex chimeric libraries containing multiple mosaicisms, high resolution recombination between closely spaced parental differences and no detectable non-chimeric parental clones. A single round of the present method, using only two parental *dszC* genes, produced chimeric genes encoding polypeptide having a rate of reaction on high-sulfur diesel increased two-fold, and a two fold increase in the amount of sulfur oxidized from a deeply biotreated diesel. The low number of null phenotypes observed in the chimeric library allowed isolates with significant improvements to be identified after screening relatively few clones. Application of a low stringency selection further identified improvements in affinity beyond that of the best parent. In addition, the high rate inherent to the other parent was captured in many of these same high affinity mutants. This significant advance in biodesulfurization has important implications in diverse fields requiring the directed evolution of proteins.

In one embodiment of the present invention, a chimeric polynucleotide is generated wherein one or more characteristics of said molecule is altered in comparison to at least one reference polynucleotide. The alteration of the chimeric polynucleotide can comprise nucleotide change, and/or amino acid changes in the encoded polypeptide, in comparison to said reference polynucleotide, polypeptide or fragment thereof. The reference polynucleotide, polypeptide or fragment thereof can be the template or fragment, or can be a molecule related to said template used for comparison. For example, where the template is a non-functional version of a polynucleotide of interest or polypeptide encoded therein, then a reference molecule can be used for comparison to chimeric polynucleotides generated. The reference molecule can be a family member of the gene or gene product of interest, such as a homologous gene, or fragment thereof. In particular embodiments, two, three or four molecules can be utilized. One of skill in the art can readily choose a reference molecule based on the templates and oligonucleotides of interest used to generate the chimeric polynucleotides.

The characteristics to be altered according to the present invention include, but are not limited to, structural motif, stability, half-life, enzyme activity, enzyme specificity, binding affinity, binding specificity, toxicity, antigenicity, interaction with an organism and interaction with components of an organism of said polynucleotide or said encoded polypeptide. A functional characteristic can be altered according to the present invention such that the activity of said functional characteristic is enhanced at a higher or lower temperature compared to a reference molecule. Furthermore, said functional activities can be enhanced in various physical or chemical environments as described above. Methods for measuring and selecting or screening these characteristics are well known in the art.

Structural motifs for proteins include, for example, alpha helices, beta sheets, solvent exposed loops, leucine zippers, α, β-barrel scaffolds and the like. Structural motifs for nucleic acids include, for example, quadraplex, aDNA, bDNA, zDNA, triple helices, stem loop, hairpin, protein binding sites and the like. Examples of regions are provided above. Methods for determining these motifs are well known in the art, for example, by nucleic acid or protein sequencing. In one

embodiment, alteration of said characteristic comprises an enhancement of said characteristic. In another embodiment, alteration of said characteristic comprises a reduction in said characteristic.

In one embodiment of the present invention, the chimera is cloned prior to selection or screening. Methods of cloning polynucleotides are well known in the art. In another embodiment, said contiguous chimera is selected or screened *in vitro* or *in vivo* prior to cloning.

The present invention allows the generation of at least one chimeric polynucleotide based in part on hybridization with at least one template. Said chimeric polynucleotides are different from said templates. Based in the method of the present invention, the differences can include, for example, an additional region, wherein said region is not present in said template. The additional region can be derived from an existing source of polynucleotides, or altered form thereof or can be a completely random sequence.

Said additional region or regions can be present at either terminus of said chimeric polynucleotide or can be present within said chimeric polynucleotide. Thus, the chimeric polynucleotide of the present invention can be longer than the template. In another embodiment, the chimeric polynucleotide can comprise an altered version of a region that is present in said template. Said region can be the same length as the region in the hybridization template or can be longer or shorter than the region in the hybridization template. Thus, the chimeric polynucleotide can be the same size, longer or shorter than the template.

In another aspect, the invention is directed to kits containing the various components necessary for performing the RACHITT™ method. Such kits can contain all the components necessary to conduct the RACHITT™ method or any portion of the components necessary to conduct the RACHITT™ method. Such kits can include components including enzymes for fragmenting oligonucleotides, for filling-in gaps, for trimming flaps, for proofreading, for incorporating uracils in templates and for modifying templates.

In one embodiment, a kit can contain a single-stranded exonuclease, *e.g.* lambda exonuclease and a non-specific endonuclease, *e.g.*, DNase I. In an alternate

embodiment, a kit can contain a single-stranded exonuclease, *e.g.* lambda exonuclease, dUTP and UNG. Such kits can also include additional components including any of the components previously listed and a polymerase, *e.g.*, T4 DNA polymerase, Taq DNA polymerase or Pfu polymerase, an endonuclease *e.g.*, Rnase HI, Taq DNA polymerase or Exonuclease VII, a ligase, *e.g.* Taq DNA ligase or T4 DNA ligase, a restriction endonuclease, *e.g.*, Pst I, Kpn I, Eco RI, Hind III, Sma I or Pvu II. A kit can be provided with instructions teaching the purchaser methods for performing the RACHITT™ method generally and in conjunction with specific polynucleotides. Typically, components of the kits are contained in separate containers. Alternatively, components utilized in sequential reactions can be contained in separate containers.

The invention will be further illustrated by the following non-limiting examples.

EXEMPLIFICATION

I. GREEN FLUORESCENCE PROTEIN CHIMERAS

Preparation of template:

One strand of the green fluorescent protein gene, GFP, was modified to contain two mutations and used as the template. The first mutation was created by cleaving the commercially available plasmid, pGFP (CLONTECH) with *XbaI*, filling in the overhanging ends using the Klenow fragment of *E. coli* polymerase I and auto-ligating to generate a four base insertion. This 4 base insertion renders the gene inactive. Conditions for the fill-in reaction were essentially as published (Molecular Cloning, Sambrook, *et al*). The ligation reaction was transformed into electrocompetent *E. coli* ElectroMax DH10B (Life Technologies) following the manufacturer's recommended protocol.

The second mutation was created by cleaving the resultant *XbaI* mutant plasmid from the above reaction with *NcoI*, filling in the overhanging ends using the Klenow fragment of *E. coli* polymerase I and auto-ligating, creating another four base insertion that would alone render the gene inactive. dUTP was then

incorporated into the target using PCR (PCR Protocols, Academic Press, Inc) using standard methods known in the art. The template was then made single stranded by digestion of one strand using lambda exonuclease (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience). Briefly, lambda exonuclease will only digest strands having a 5' phosphate. Therefore, to generate single stranded DNA, the polynucleotide is amplified with one primer containing a 5' phosphate and the other primer not containing a 5' phosphate.

Preparation and hybridization of hybridizing fragments:

Oligonucleotides for use in hybridizing the above-described GFP template to generate chimeric polynucleotides were chemically synthesized oligonucleotides from a commercial source. These fragments were synthesized with specific mutations with respect to the template. Since the full-length template above was rendered single stranded, the synthetic oligonucleotides were designed to anneal the single stranded template. Some internal oligonucleotides contained a 5'-phosphate (PO_4) modification for efficient ligation with fragments hybridized upstream on the template. The following three custom synthesized oligonucleotides (Sigma-Genosys) were designed as three of the four fragments of interest used on the above double mutant GFP template. Fragment 2 incorporated a four base pair deletion with respect to the template:

Chemically synthesized fragments:

Fragment 1:

5'-CACAGGAAACAGCTATGACC (SEQ ID NO. 1)

Fragment 2:

5'-PO_4-GCATGCCTGCAGGTCGACTCTAG*AGGATCCCCGGGTACCGGTAG AAAAAATGAGTAAAGGAG (SEQ ID NO. 2)

The location of the above deletion is marked by the *.

Fragment 4:
5'-PO_4-CGCTACCATTACCAGTTGGTCTGGTG (SEQ ID NO. 4)

Fragment 3 was generated by PCR of the wild type GFP gene using upstream and downstream primers as follows. An upstream PCR primer that binds between the two mutations was designed to have a 5'-tail (bold type) of sequence unrelated to the template:

Upstream primer for Fragment 3:
5'-TTATATACTAGTATCGATGATCGAGGAGGATGATTAAATGAGTAAA GGAGAAGAACTTTTC (SEQ ID NO. 5)

Downstream primer for Fragment 3:
5'-CACCAGACCAACTGGTAATGGTAGC (SEQ ID NO. 6)

The fragment generated by PCR using the fragment 3 primers was rendered single stranded as above and all four fragments were hybridized to the transient template. Fragment 3 has the sequence

(TTATATACTAGTATCGATGATCGAGGAGGATGATTAAATGAGTAAAGG AGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTG ATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGC AACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTA CCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATG CTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTTTCAAGAGT GCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCT TGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAAC ATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACA TCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAG ACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAA

AATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAGCATTCGTAGAATTCCAACTGAGCGCCGGTCGCTACCATTACCAGTTGGACTGGTG). SEQ ID NO.: 7.

A schematic diagram of the hybridization of fragments 1-4 to the template is shown in Figure 2. A double mutant, single stranded transient template (TT: bottom panel, X marks the sites of mutation) was derived from a double mutant GFP gene (top panel) as described above. The hybridized fragments (bottom panel) are numbered 1 to 4.

II

In another example, oligonucleotides were generated from the wild type GFP gene. The GFP gene was amplified by PCR, rendered single stranded by lambda exonuclease cleavage, randomly cleaved using DNaseI (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience) and hybridized to the transient template as above. Before hybridization, the DNaseI products ranging in size from 40 to over 200 bases were isolated from a 6 M urea-PAGE gel, eluted and purified using standard procedures (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience).

III

In a third example, no wild type oligonucleotides of interest were used. Two GFP mutants that were singly mutated in either the *XbaI* or *NcoI* sites (by the above procedure) were simultaneously used as sources of oligonucleotides. Randomly cleaved oligonucleotides of each single mutant gene of interest were generated as above and hybridized to the double mutant template.

Hybridization:

7 to 35 fmoles of template and 240 fmoles to 5 pmoles of oligonucleotides produced according to I, II and III above were incubated in 1X *Taq* ligase buffer

with 0.1 μg *E. coli* tRNA in a total volume of 10 μl to allow hybridization of oligonucleotides to regions of template having sufficient homology.

Trimming:

The hybridized oligonucleotides were trimmed such that non-hybridized termini were removed. The trimming reaction for 5' flaps (or unhybridized termini) employed the flap endonuclease activity of *Taq* DNA polymerase, as described (Lyamichev, *et al, Science 260*: 778-783, 1993) or for both 5' and 3' termini, Exonuclease VII (Boehringer Mannheim) according to the manufacturer's recommendation. The trimming reaction using *Taq* polymerase was stopped by cooling on ice, while the exonuclease VII reaction was stopped by heating to 75°C for 15 minutes.

Gap filling:

Gaps between hybridized oligonucleotides were then filled using T4 DNA polymerase (New England Biolabs), according to the manufacturer's recommendation, or 0.5 units Red*Taq* DNA polymerase (Sigma), while ramping the temperature from 45 to 60 degrees over 30 minutes. The gap filling reaction using *Taq* polymerase was stopped by cooling on ice. The T4 DNA polymerase gap filling reaction was stopped by heating to 75°C, for 15 min.

Ligation:

Ligation of hybridized oligonucleotides was carried out using *Taq* DNA ligase (New England Biolabs) and/or T4 DNA ligase (New England Biolabs) according to the manufacturers' recommendations. All reactions at or below 30°C proceeded in New England Biolabs buffer II, supplemented with 1mM ATP. All reactions above 30°C proceeded in New England Biolabs *Taq* ligase buffer. After ligation, 0.5 U uracil DNA glycosylase (UDG) (New England Biolabs) was added and the sample was incubated at 37°C for 30 minutes in 1X UDG buffer. This treatment rendered the template unusable as a substrate in PCR. PCR was then used to amplify the newly synthesized chimera and its complementary natural base containing strand according to Table I.

Table I

Volume (µl)	
2.5	UDG digested chimeric population
31.25	H_2O
5	Red *Taq* Buffer, 10x
4	2.5 mM dNTPs
3	5 µM upstream primer SEQ ID NO: 5
3	5 µM downstream primer SEQ ID NO: 6
1.25	Red *Taq* polymerase (1U/µl)
Cycling Parameters	
94°C, 3 minutes	
94°C, 30 seconds	25 cycles
55°C, 20 seconds	
72°C, 45 seconds	
72°C, 10 minutes	

Molecular cloning of the PCR products into the original pGFP vector backbone was performed after purification using UltraClean Mini Plasmid Prep Kit cartridges (MoBio) and restriction digestion of plasmid and insert with *EcoRI* and *HindIII* to yield flanking sticky ends. Each of these steps is within the abilities of one skilled in the art and are described in detail in common methods manuals (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience).

In all the examples above, chimeric products were observed using restriction endonuclease digestion and analysis of cleavage products by agarose gel electrophoresis and/or by cloning the products and observing the presence or

absence of fluorescence in the corresponding expressed polypeptide, the green fluorescent protein.

Chimeric polynucleotide formation was assayed by restriction analysis across several experiments. Chimeric strands incorporating at least two annealing oligonucleotides comprised approximately 40 to 60% of the total products formed, whereas no detectable background cleavage was observed when oligonucleotides were omitted from the reaction. Screening for restoration of the fluorescent phenotype revealed that approximately 14% of the resulting products had acquired both altered loci by incorporation of at least three annealing oligonucleotides. Omission of oligonucleotides or inclusion of oligonucleotides from just one singly-mutated parental gene, produces no fluorescent colonies.

II. CHIMERAGENESIS OF THE *dszC* GENE:

Preparation of template:

The transient template was one strand of the DBT monooxygenase gene, *dszC*, from the soil microorganism, *Rhodococcus erythropolis* IGTS8. dUTP was incorporated into the target using PCR (PCR Protocols, Academic Press, Inc), and the template was made single stranded by digestion of one strand using lambda exonuclease (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience).

Preparation and hybridization of hybridizing fragments:

The wild type *dszC* gene from the free living microorganism, *Nocardia sp.* A3H1 was amplified by PCR, made single stranded as described above and randomly cleaved using DNaseI (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience) and then hybridized to the transient template as described above. Before hybridization, the DNaseI products ranging in size from 40 to 200 bases were isolated from a 6 M urea-PAGE gel and purified using standard procedures (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience).

Before the above procedures, both template and hybridizing fragments were subjected to PCR amplification (as described above) using the following primers, which included 5' phosphorylations, where appropriate:

Forward primer 1:

5'- CCGTCGCCCTCGATCAGTG (SEQ ID NO: 20)

Reverse primer 1:

5'-GGGTGGTGTGGCTGGAAAC (SEQ ID NO: 21)

In addition to the randomly cleaved fragments of the A3H1 genes, two synthetic oligonucleotides were included to anneal to the 5' and 3' termini of the template, respectively. The latter oligonucleotide included a 5' phosphate as above. The 5' binding oligo was Forward primer 1, the second was:

3'-binding oligo 1:

5'-PO4-CGCTACCATTACCAGTTGGTCTGGTG (SEQ ID NO: 22)

Trimming:

Two methods were successfully used in making *dszC* chimera. In the first, Exonuclease VII was used to trim 5' and 3' flaps and was used according to the manufacturer's recommendation. The Exonuclease VII reaction was stopped by heating to 75°C, 15min. In the second method, the 5' flap endonuclease activity of *Taq* DNA polymerase was used as described (Lyamichev). This reaction was stopped by cooling on ice. Where *Taq* DNA polymerase was used, Pfu DNA polymerase was included in the gap filling step, below in order to trim 3' flaps.

Gap filling:

Gaps remaining on the template were filled using T4 DNA polymerase at 12°C to 37°C(New England Biolabs), according to the manufacturers recommendation, or 0.5 units Red*Taq* DNA polymerase (Sigma) and 1.25 units Pfu

DNA polymerase (Stratagene), while ramping the temperature from 45 to 60 degrees over 30 minutes. The Red*Taq* DNA polymerase gap filling reaction using *Taq* polymerase was stopped by cooling on ice. The T4 DNA polymerase gap filling reaction was stopped by heating to 75°C, 15 min.

Ligation:

Ligation was carried out using *Taq* DNA ligase (New England Biolabs) and/or T4 DNA ligase (New England Biolabs) according to the manufacturers' recommendations.

After ligation, 0.5 U uracil DNA glycosylase (UDG) (New England Biolabs) was used 37°C, 30 min to render the template unusable as a substrate in the PCR. PCR then was used to amplify the newly synthesized product strand and its complementary, chimeric, natural base containing strand as follows.

Table II

Volume (µl)	
2.5	UDG digested chimeric population
31.25	H_2O
5	Red *Taq* Buffer, 10x
4	2.5 mM TPs
3	5 µM forward primer, SEQ ID NO: 20
3	5 µM reverse primer, SEQ ID NO: 21
1.25	Red *Taq* polymerase (1 U/µl)
Cycling Parameters	
94°C, 3 minutes	
94°C, 30 seconds	
62°C, 30 seconds	30 cycles
72°C, 120 seconds	
72°C, 10 minutes	

Molecular cloning of the PCR products into an appropriate shuttle vector backbone was performed after purification using UltraClean Mini Plasmid Prep Kit cartridges (MoBio) and restriction digestion of plasmid and insert with *XbaI* and *SfiI* to yield flanking sticky ends. Each of these steps is within the abilities of one skilled in the art and are described in detail in common methods manuals (Molecular Cloning, Second Edition, Cold Spring Harbor Laboratory Press; or Current Protocols, Wiley Interscience). In the examples above, chimeric products were observed using restriction endonuclease digestion and analysis of cleavage products by agarose gel electrophoresis and/or by cloning the products and observing the presence or absence of improvements in the corresponding expressed chimeric DBT monooxygenase.

Results of chimeric *dszC* library:

The *dszC* chimeric library is highly random:

Clones from the chimeric *dszC* library were tested for the presence of chimeric sequences by restriction endonuclease cleavage using four sites along the gene (*BglI, XmnI, XmaI/SmaI and BglII*). The corresponding enzyme will either cleave or not cleave, depending upon which parental gene donated the sequence at that site. Every theoretically possible chimera analyzed by this test will have some permutation of cleavage or no cleavage at each of the four positions. There are thus 16 possible classes, each having a unique permutation of one or the other parental sequences present at each of the four sites. To date, 38 clones have been tested in this fashion. The results of these 38 assays are shown in Table III, which shows the parental contribution of each clone at each site. The clones are grouped by similar patterns of cleavage. Of the 16 possible classes, representatives have been observed for 14 classes. For comparison, for independent, random distribution at each site, a statistical analysis predicts 14 +/- 1, making the pattern observed statistically indistinguishable from random to well within one standard deviation. Remarkably, only eight of these 38 isolates were actually chosen at random. The others were chosen because of their growth on selection plates, after assessment of their ability

to effect indole conversion or after confirmation of higher activity in a microtiter format assay.

For indole conversion, as is the case with many oxygenases, the DBT-MO catalyzes the conversion of indole to the colored product, indigo. We found that colonies containing the less active (on Cx-DBT) *dszC* gene from A3H1 produce consistently and significantly more indigo than does the more active IGTS8 *dszC* gene. This discrepancy suggested specificity differences of the two parents for ratios of indole:Cx-DBT conversion. It was anticipated that the chimeric library should contain *dszC* clones with a wide range of indole oxidase activity. This was indeed the case, with colonies ranging in activity from having no visible production of indigo, less indigo oxidase activity than the less active (on indigo) IGTS8 *dszC*, activity similar to that enzyme, intermediate between the two parents, activity equal to the more active A3H1 DBT-MO and one that produced more indigo than the best parent (enough to excrete indigo from the colony into the plate, which was not observed with either parent). Quantitative assays revealed chimera that were 20-fold improved over the parental enzymes. Supporting the assertion that these were true specificity differences and not mere differences in cellular activity (*e.g.*, expression levels), it was found after testing that, just as in the parents, the DBT-MO activity in the chimeric clones did not correlate with the level of indigo production.

For the microtiter assay for DBT-MO activity, the biodesulfurization pathway genes, *dszABCD*, encode the proteins responsible for the release of sulfur from alkylated dibenzothiophenes in fossil fuels. Since the *dszC* product, the dibenzothiophene monooxygenase, DBT-MO, is rate limiting in one host strain, we used the overall conversion of the dibenzothiophenic sulfur in diesel fuel to sulfite as a measure of improvements in DBT-MO and therefore as a measure of the evolution of *dszC*. The assay consisted of shaking cells in aqueous growth medium with diesel fuel in an organic phase and then spectrophotometrically measuring the sulfite produced in a given amount of time.

For growth on selection plates, the chimeric *dszC* library clones were placed on solid medium containing diesel fuel (25 ppm final sulfur) as the sole sulfur

source. While cells lacking *dszC* failed to grow on this medium, cells containing an intact *dszC*, including thousands of chimeric *dszC*s, were able to release the sulfur from the oil and grew normally.

TABLE III

Restriction Enzymes:					
Class:	BglI	XmnI	XmaI/SmaI	BglII	number of clones per class (n=38)
1	IGTS8	IGTS8	IGTS8	IGTS8	1
2	A3H1	IGTS8	IGTS8	IGTS8	1
3	IGTS8	A3HI	IGTS8	IGTS8	2
4	IGTS8	IGTS8	A3HI	IGTS8	1
5	IGTS8	IGTS8	IGTS8	A3HI	0
6	A3HI	A3HI	IGTS8	IGTS8	0
7	A3HI	IGTS8	A3HI	IGTS8	7
8	A3HI	IGTS8	IGTS8	A3HI	2
9	IGTS8	A3HI	A3HI	A3HI	1
10	IGTS8	IGTS8	A3HI	A3HI	3
10'	IGTS8	IGTS8	A3HI+ New Site	A3HI	1
11	IGTS8	A3HI	IGTS8	A3HI	1
12	A3HI	A3HI	A3HI	IGTS8	5
13	A3HI	A3HI	IGTS8	A3HI	6
14	A3HI	IGTS8	A3HI	A3HI	4
15	IGTS8	A3HI	A3HI	A3HI	2
16	A3HI	A3HI	A3HI	A3HI	1

DNA sequence of *dszC* chimeras:

Several chimeric *dszC* genes from the above library were prepared, purified and subjected to DNA sequencing by standard methods. Templates chosen included both selected and unselected chimeric clones. With the first data interpreted (one sequencing reaction of about 400 out of a total of 1248 nucleotides for each of 9

isolates), it was clear that the present invention generated a wide number of diverse chimeras. The number of A3H1 hybridizing fragments in the analyzed 400 bases ranged from 1 to 4 fragments, indicating that some chimeric genes may have eight or more sections from the A3H1 parent (see below). The resolution of recombination from the two parents appears to be remarkably fine, with several cross-overs occurring to separate or bring together parental differences within stretches of less than 5 nucleotides in length. This is a higher frequency of fine resolution recombination than published for other methods and is expected to result in more diverse libraries that in general will contain improved clones that are not available by other methods.

Figure 10 shows the completed DNA sequence of one evolved, chimeric *dszC* (SEQ ID. NO: 3). It is aligned with the *Nocardia* sp. A3H1 parental *dszC* above (SEQ ID NO: 23)and the IGTS8 *dszC* (SEQ ID NO: 24) below, all shown with the entire reading frame from the initial start codon to the final stop codon. The locations of three introduced mutations in the chimeric *dszC*, but not acquired from either parent, are indicated by triangles below the sequences. Where the two parental *dszC* genes differ, the chimeric *dszC* sequence is highlighted in black to show identity to A3H1, or by underlines at individual nucleotides for IGTS8. Where the two parents and the chimeric *dszC* are identical, an asterisk was placed below the sequence alignment. The longer underlined sequences that connect the black highlighted A3H1 donated differences that have no intervening IGTS8 differences and may or may not correspond to regions donated by a single A3H1 hybridizing fragment. The single underlined regions delineate the transition from sequences donated by one or the other parent and indicate the maximum possible length of each hybridizing fragment.

On average, the hybridizing fragments are expected to extend to about the middle of each underlined region. This alignment makes immediately apparent several characteristics of the present method for chimeragenesis. For example, the templates and hybridizing fragments were designed to allow either parent to donate sequences all the way to the ends of the gene. In this case, the hybridizing fragments contributed both ends. (In the above mentioned DNA sequences analyzed

so far from the beginning of the gene, both parents have contributed.) It was also anticipated that insertions or deletions could be introduced on hybridizing fragments, and that is the case as seen for the introduction of a 3 nucleotide deletion in line one (dashed lines). An 8 nucleotide insertion carried on hybridizing fragments was also readily incorporated in two other sequences (not shown) and much larger insertions or deletions should be possible. The fine resolution of recombination mentioned in the discussion of the first third of the chimeric *dszCs* was observed here throughout the gene, with at least five switches between A3H1 and IGTS8 derived sequences occurring between mutational differences lying 5 or fewer sequence identities apart. The number of hybridizing fragments is also quite high, with at least ten binding to the template that gave rise to this one isolate. The longest A3H1-boxed fragment (extending from line 6 to 8 in the figure) represents hybridizing fragment(s) of 218 to 289 nucleotides.

Representing the other extreme, the smallest number of A3H1-donated sequence differences (see line 2) is only 1 mutation long. The most likely explanation here is that a hybridizing fragment of less than 92 nucleotides (the size of the flanking underlined region) was incorporated or that a longer hybridizing fragment bound, but was trimmed back to within the smaller underlined region of sequence identity. A complementary feature of this sequence alignment is the lone IGTS8 difference (double underlined) in the middle of line four. The upstream hybridizing fragment ended in the single underlined text to the left of this nucleotide, while the next downstream hybridizing fragment started in the single underlined region to its right. Taken together, these last two observations are clear demonstration that this method, which relies on the binding of fragments large enough to hybridize despite mismatches with respect to the template scaffold, can nevertheless recombine single nucleotide differences between parental genes.

In total, of 127 nucleotide differences between the two parents, in the particular chimeric clone shown in Figure 10, 49 nucleotide differences were acquired from the IGTS8 template and 78 came from A3H1. This number can be adjusted easily by altering the ratio of hybridizing oligonucleotides to template molecules.

Demonstration of improvements in the DBT monooxygenase due to the chimeric genes in the chimeric *dszC* library:

From the above library, 367 clones were analyzed in a microtiter-format assay described above for activity. Similar assays are common and within the ability of the skilled artisan. This assay revealed seven clones that were from 140% to 330% faster at releasing sulfur from a 20% diesel fuel feedstock than the fastest of the two parental genes (Figure 9).

Growth selection of the chimeric *dszC* library:

In a 10 day growth selection, the chimeric *dszC* library (in the presence of the other *dsz* genes, *dszA*, *dszB* and *dszD*, and in an appropriate *Rhodococcus* host strain), allowed growth and desulfurization on a diluted, biotreated diesel fuel to about the same extent as one of the library parents, the A3H1 *dszC*. The library's other parent, *dszC* from IGTS8, while much more active on undiluted, unbiotreated distillates, has little activity on this substrate and does not allow significant growth of the same host strain. The microtiter analysis mentioned above indicated that many chimeras of the library that had the growth properties of the one parent also had acquired the high activity unique to the other parent. The recombination of two parental phenotypes into one progeny gene is a powerful demonstration of a valid method for chimeragenesis.

In order to confirm that these new bifunctional clones were indeed chimeric, several were subjected to the analysis mentioned above. It was found that all the bifunctional clones were indeed chimeric (see Table IV)

TABLE IV

Restriction Enzymes:				
	BglI	XmnI	XmaI/SmaI	BglII
1	A3HI	IGTS8	IGTS8	IGTS8
2	IGTS8	IGTS8	A3HI	IGTS8
3	IGTS8	IGTS8	A3HI	A3HI
4	A3HI	A3HI	A3HI	IGTS8
5	A3HI	A3HI	A3HI	IGTS8
6	A3HI	A3HI	IGTS8	A3HI
7	A3HI	A3HI	IGTS8	A3HI
8	A3HI	IGTS8	A3HI	A3HI
9	IGTS8	A3HI	A3HI	A3HI

The teachings of all references, patents and patent applications cited herein are hereby incorporated by reference in their entireties. While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A method for forming at least one chimeric polynucleotide comprising;

   contacting a single-stranded polynucleotide template with a random population of oligonucleotides, under conditions wherein at least two oligonucleotides hybridize to the template; and

   treating the hybridized oligonucleotides such that a chimeric polynucleotide is formed.

2. The method of Claim 1, wherein the population of oligonucleotides randomly hybridize to the template.

3. The method of Claim 1, wherein the oligonucleotides hybridize to the template under conditions of low stringency.

4. The method of Claim 1, further comprising removing a step of trimming flaps and/or loops.

5. The method of Claim 4, wherein 5′ flaps are trimmed using an enzyme selected from the group consisting of Rnase HI, *Taq* DNA polymerase, flap endonuclease or Exonuclease VII.

6. The method of Claim 1, further comprising a step of filling in gaps.

7. The method of Claim 4, wherein the flaps are removed using an enzyme selected from the group consisting of: RNase HI, *Taq* DNA polymerase, flap endonuclease and Exonuclease VII.

8. The method of Claim 4, wherein 3' flaps are removed and gaps are filled using an enzyme selected from the group consisting of *Pfu* polymerase, *Taq* DNA polymerase, DNA pol I and Klenow.

9. The method of Claim 1, wherein the hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

10. The method of Claim 1, wherein the population of oligonucleotides comprises oligonucleotides generated by *in vitro* amplification or synthetic production.

11. The method of Claim 1, further comprising repeating steps of the first method a second time using the chimeric polynucleotide formed in the first method to generate the template or the random population of oligonucleotides.

12. The method of Claim 1, further comprising selectively amplifying the chimeric polynucleotide with respect to the template.

13. The method of Claim 12, wherein the template is selectively removed or destroyed.

14. The method of Claim 12, wherein the template includes at least one uracil.

15. The method of Claim 1, further comprising cloning at least one chimeric polynucleotide.

16. The method of Claim 1, further comprising expressing at least one chimeric polynucleotide.

17. The method of Claim 1, further comprising screening or selecting a chimeric polynucleotide having at least one desired characteristic.

18. The method of Claim 1, wherein the template and at least one oligonucleotide encode a desulfurizing enzyme or fragment thereof.

19. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 9, an altered version or homolog thereof.

20. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 11, an altered version or homolog thereof.

21. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 13, an altered version or homolog thereof.

22. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 14, an altered version or homolog thereof.

23. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 17, an altered version or homolog thereof.

24. The method of Claim 18, wherein the desulfurizing enzyme comprises SEQ ID NO: 19, an altered version or homolog thereof.

25. The method of Claim 1, wherein at least one oligonucleotide comprises a region of random sequence.

26. The method of Claim 1, wherein the template comprises at least one region of random sequence.

27. A method for forming a chimeric polynucleotide comprising;

contacting a single-stranded polynucleotide template with a population of oligonucleotides under conditions such that at least two oligonucleotides hybridize to a given template, and wherein the population of oligonucleotides comprises oligonucleotides such that two or more regions of the template are complementary to two or more oligonucleotides of the population; and

ligating the hybridized oligonucleotides such that one chimeric polynucleotide, is generated.

28. The method of Claim 27, wherein the oligonucleotides hybridize to the template under conditions of low stringency.

29. The method of Claim 27, wherein the oligonucleotides comprise a random population of oligonucleotides.

30. The method of Claim 27, further comprising a step of removing flaps or loops.

31. The method of Claim 30, wherein the flaps are removed using an enzyme selected from the group consisting of RNase HI, *Taq* DNA polymerase or Exonuclease VII.

32. The method of Claim 27, further comprising a step of filling in gaps.

33. The method of Claim 32, wherein gaps are filled using an enzyme selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase.

34. The method of Claim 32, wherein 3' flaps are removed and gaps are filled using an enzyme selected from the group consisting of *Pfu* polymerase, T4 DNA polymerase, DNA polI and Klenow.

35. The method of Claim 27, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

36. The method of Claim 27, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

37. The method of Claim 27, further comprising repeating the steps of the first method a second time using the chimeric polynucleotide formed in the first method to generate the template or the random population of oligonucleotides.

38. The method of Claim 27, further comprising selectively amplifying the chimeric polynucleotide with respect to the template.

39. The method of Claim 38, wherein the templates are selectively removed or destroyed.

40. The method of Claim 38, wherein the template includes at least one uracil.

41. The method of Claim 27, further comprising cloning at least one chimeric polynucleotide.

42. The method of Claim 27, further comprising expressing at least one chimeric polynucleotide.

43. The method of Claim 27, further comprising screening or selecting a chimeric polynucleotide having at least one desired characteristic.

44. The method of Claim 27, wherein the template and at least one of the oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

45. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 9, an altered version or homolog thereof.

46. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 11, an altered version or homolog thereof.

47. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 13, an altered version or homolog thereof.

48. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 15, an altered version or homolog thereof.

49. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 17, an altered version or homolog thereof.

50. The method of Claim 44, wherein the desulfurizing enzyme comprises SEQ ID NO: 19, an altered version or homolog thereof.

51. The method of Claim 27, wherein the template comprises at least one region of random sequence.

52. The method of Claim 27, wherein at least one oligonucleotide comprises a region of random sequence.

53. A method for forming at least one chimeric polynucleotide comprising:

contacting a single-stranded polynucleotide template with a population of oligonucleotides, wherein at least two of the oligonucleotides hybridize to the same template, such that at least one flap is formed;

removing at least one flap; and

ligating the hybridized oligonucleotides such that one chimeric polynucleotide, is generated.

54. The method of Claim 53, wherein the oligonucleotides hybridize to the template under conditions of low stringency.

55. The method of Claim 53, wherein the oligonucleotides comprise a random population of oligonucleotides.

56. The method of Claim 53, wherein flaps are removed using RNase HI, flap endonuclease or Exonuclease VII.

57. The method of Claim 53, further comprising filling in gaps.

58. The method of Claim 57, wherein gaps are filled using an enzyme selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

59. The method of Claim 57, wherein flaps removed and gaps are filled using an enzyme selected from the group consisting of *Pfu* polymerase, T4 DNA polymerase, DNA pol I or Klenow.

60. The method of Claim 53, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

61. The method of Claim 53, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

62. The method of Claim 53, further comprising repeating the steps of the first method a second time using the chimeric polynucleotide formed in the first method to generate the template or the random population of oligonucleotides.

63. The method of Claim 53, further comprising selectively amplifying the chimeric polypeptide with respect to the templates.

64. The method of Claim 63, wherein the template is selectively removed or destroyed.

65. The method of Claim 63, wherein the template includes at least one uracil.

66. The method of Claim 53, further comprising cloning at least one chimeric polynucleotide.

67. The method of Claim 53, further comprising expressing at least one chimeric polynucleotide.

68. The method of Claim 53, further comprising screening or selecting at least one chimeric polynucleotide having desired characteristics.

69. The method of Claim 53, wherein the template and at least one of the oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

70. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 9, an altered version or homolog thereof.

71. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 11, an altered version or homolog thereof.

72. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 13, an altered version or homolog thereof.

73. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 15, an altered version or homolog thereof.

74. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 17, an altered version or homolog thereof.

75. The method of Claim 69, wherein the desulfurizing enzyme comprises SEQ ID NO: 19, an altered version or homolog thereof.

76. The method of Claim 53, wherein the template comprises at least one region of random sequence.

77. The method of Claim 53, wherein at least one oligonucleotide comprises a region of random sequence.

78. A method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide, comprising the steps of:

contacting a least one single-stranded polynucleotide template with a random population of oligonucleotides under conditions wherein at least two oligonucleotides hybridize to the template;

filling in gaps between adjacently hybridized oligonucleotides;

ligating hybridized oligonucleotides to form one chimeric polynucleotide hybridized to a template;

selectively amplifying the chimeric polynucleotide with respect to the templates; and

selecting or screening at least one chimeric polynucleotide, wherein the specified characteristic is altered in comparison to the reference polynucleotide.

79. The method of Claim 78, wherein the oligonucleotides hybridize to the template under conditions of low stringency.

80. The method of Claim 78, further comprising removing flaps or loops.

81. The method of Claim 78, wherein flaps are removed using an enzyme selected from the group consisting of RNase HI, flap endonuclease and Exonuclease VII.

82. The method of Claim 78, wherein gaps are filled using an enzyme selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and Pfu polymerase.

83. The method of Claim 78, wherein flaps are removed and gaps are filled using an enzyme selected from the group consisting of: *Pfu* polymerase T4 DNA polymerase, DNA pol I and Klenow.

84. The method of Claim 78, wherein the hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

85. The method of Claim 78, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

86. The method of Claim 78, wherein at least one oligonucleotide comprises a region of random sequence.

87. The method of Claim 78, wherein the template includes at least one uracil.

88. The method of Claim 78, further comprising repeating the steps of the first method a second time using the chimeric polynucleotide formed in the first method to generate the template or the random population of oligonucleotides.

89. The method of Claim 78, wherein the template and at least one of the oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

90. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 9, an altered version or homolog thereof.

91. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 11, an altered version or homolog thereof.

92. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 13, an altered version or homolog thereof.

93. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 15, an altered version or homolog thereof.

94. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 17, an altered version or homolog thereof.

95. The method of Claim 89, wherein the desulfurizing enzyme comprises SEQ ID NO: 19, an altered version or homolog thereof.

96. The method of Claim 78, wherein the template comprises at least one region of random sequence.

97. The method of Claim 78, wherein the template is selectively removed or destroyed.

98. The method of Claim 97, wherein the template comprises at least one uracil.

99. The method of Claim 78, further comprising cloning at least one chimeric polynucleotide.

100. The method of Claim 78, further comprising expressing at least one chimeric polynucleotide.

101. A method for forming a chimeric polynucleotide comprising:

contacting a single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

filling in gaps between the hybridized oligonucleotides; and

ligating the hybridized oligonucleotides such that a chimeric polynucleotide is formed.

102. The method of Claim 101, wherein the population of oligonucleotides comprise a random population of oligonucleotides.

103. The method of Claim 101, further comprising a step of trimming flaps.

104. The method of Claim 103, wherein both 3' and 5' flaps are trimmed.

105. The method of Claim 101, wherein the template comprises a coding region.

106. The method of Claim 105, wherein a template-length chimeric polynucleotide is formed.

107. The method of Claim 101, wherein the population of oligonucleotides is produced by fragmenting a single-stranded nucleic acid.

108. The method of Claim 101, wherein the population of oligonucleotides is produced synthetically.

109. The method of Claim 101, wherein a plurality of chimeric polynucleotides is formed.

110. The method of Claim 109, wherein the number of chimeric polynucleotides formed and the number of single-stranded templates is in a ratio of about 1.

111. A non-reiterative process for forming a template-length chimeric polynucleotide comprising the following steps:

contacting a single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and

ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide.

112. The method of Claim 111, wherein the population of oligonucleotides comprise a random population of oligonucleotides.

113. The method of Claim 111, further comprising a step of filling-in gaps.

114. The method of Claim 113, further comprising a step of trimming flaps.

115. The method of Claim 114, wherein both 3' and 5' flaps are trimmed.

116. The method of Claim 111, wherein the population of oligonucleotides is formed by fragmenting a single-stranded nucleic acid.

117. The method of Claim 111, wherein the population of oligonucleotides is formed synthetically.

118. The method of Claim 111, wherein a plurality of chimeric polynucleotides is formed.

119. The method of Claim 118, wherein the number of chimeric polynucleotides formed and the number of single-stranded templates is in a ratio of about 1.

120. A method for forming a chimeric polynucleotide comprising the following steps:

contacting a single-stranded template with a population of oligonucleotides produced by fragmenting a single-stranded nucleic acid or by chemical synthesis, under conditions such that at least two of the oligonucleotides hybridize to the template; and

ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide.

121. The method of Claim 101, wherein the population of oligonucleotides comprise a random population of oligonucleotides.

122. The method of Claim 120, further comprising a step of filling-in gaps.

123. The method of Claim 122, further comprising a step of trimming flaps.

124. The method of Claim 123, wherein both 3' and 5' flaps are trimmed.

125. The method of Claim 120, wherein a plurality of chimeric polynucleotides is formed.

126. The method of Claim 125, wherein the number of chimeric polynucleotides formed and the number of single-stranded templates is in a ratio of about 1.

127. A method for forming a plurality of chimeric polynucleotides on single-stranded polynucleotide templates, wherein the number of chimeric polynucleotides formed and the number of single-stranded templates is in a ratio of about 1 comprising the following steps:

contacting a single-stranded template with a population of oligonucleotides produced by fragmenting a single-stranded nucleic acid or by chemical synthesis, under conditions such that at least two of the oligonucleotides hybridize to the template; and

ligating the hybridized oligonucleotides, thereby forming a template-length chimeric polynucleotide.

128. The method of Claim 127, wherein the population of oligonucleotides comprise a random population of oligonucleotides.

129. The method of Claim 127, further comprising a step of filling-in gaps.

130. The method of Claim 129, further comprising a step of trimming flaps.

131. The method of Claim 130, wherein both 3' and 5' flaps are trimmed.

132. The method of Claim 127, wherein the population of oligonucleotides is formed by fragmenting a single-stranded nucleic acid.

133. The method of Claim 127, wherein the population of oligonucleotides is formed synthetically.

134. A kit for performing a method of directed evolution comprising components selected from the group of components for fragmenting oligonucleotides, for filling-in gaps, for trimming flaps, for proofreading, for incorporating uracils in templates and for modifying templates.

135. The kit of Claim 134 further comprising instructions for performing a method of directed evolution.

136. A kit for performing a method of directed evolution comprising components consisting of lambda exonuclease and DNase I.

137. The kit of Claim 36 further comprising instructions for performing a method of directed evolution.

138. The kit of Claim 137, further comprising a component selected from the group of components including those for filling-in gaps, for trimming flaps, for proofreading, for incorporating uracils in templates and for modifying templates.

139. A kit for performing a method of directed evolution comprising components consisting of lambda exonuclease, dioxyuracil triphosphate and uracil DNA glycosylase.

140. The kit of Claim 139 further comprising instructions for performing a method of directed evolution.

141. The kit of Claim 140 further comprising a component selected from the group of components including those for filling-in gaps, for trimming flaps and for proofreading.

142. A method for forming a chimeric polynucleotide comprising:

preparing a single-stranded template comprising RNA;

contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and

treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed.

143. The method of Claim 142, wherein the template comprises mRNA.

144. The method of Claim143, further comprising the step of selectively amplifying the chimeric polynucleotide relative to the mRNA template.

145. The method of Claim 144, wherein the selective amplification of the chimeric polynucleotide is achieved by an *in vitro* nucleic acid amplification.

146. The method of Claim 142, wherein treating the hybridized oligonucleotides comprises an enzymatic treatment.

147. The method of Claim 142, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* nucleic acid amplification or synthetic production.

148. The method of Claim 142, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

149. The method of Claim 142, wherein the single-stranded template comprises at least one region of random or partially random sequence.

150. The method of Claim 142, further comprising selecting a chimeric polynucleotide for a particular characteristic.

151. The method of Claim 142, further comprising at least one step selected from the group of steps consisting of trimming flaps, filling in gaps between hybridized oligonucleotides and ligating hybridized oligonucleotides.

152. The method of Claim 151, wherein the step of trimming flaps precedes the step of filling in gaps between hybridized oligonucleotides.

153. The method of Claim 151, wherein the step of trimming flaps occurs concurrently with the step of filling in gaps between hybridized oligonucleotides.

154. The method of Claim 151, wherein gaps are filled in using a polymerase with reverse transcriptase activity.

155. The method of Claim 151, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

156. The method of Claim 151, wherein the single-stranded mRNA template is protected during the step of trimming flaps.

157. The method of Claim 156, wherein the single-stranded mRNA template is protected by annealing a preselected oligonucleotide to an end of the template.

158. A method for generating a chimeric polynucleotide, wherein at least one characteristic of the chimeric molecule is altered in comparison to a reference polynucleotide, comprising the steps of:

preparing a single-stranded polynucleotide template comprising mRNA;

contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

trimming flaps;

filling in gaps between hybridized oligonucleotides;

ligating hybridized oligonucleotides to form at least one chimeric polynucleotide;

selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening the chimeric polynucleotide, wherein at least one characteristic is altered in comparison to the reference polynucleotide.

159. The method of Claim 158, wherein the single-stranded mRNA template is protected by annealing a preselected oligonucleotide to an end of the template.

160. The method of Claim 158, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

161. The method of Claim 158, wherein gaps are filled in using a polymerase with reverse transcriptase activity.

162. A chimeric polynucleotide prepared according to the method of Claim 158.

163. A method for preparing a single-stranded polynucleotide template containing mRNA suitable for use in forming a chimeric polynucleotide.

164. A method for preparing an mRNA transient template *in vivo* for use in forming a chimeric polynucleotide comprising:

preparing a DNA molecule comprising a suitable insert to be transcribed;

transforming a suitable cell with the DNA molecule such that transcription of the DNA molecule occurs;

lysing the transformed cell; and

purifying the transcribed mRNA product from the cell lysate.

165. The method of Claim 164, further comprising the step of protecting the mRNA template.

166. The template formed by the method of Claim 165.

167. A method for preparing an mRNA transient template *in vitro* for use in forming a chimeric polynucleotide comprising:

preparing a DNA molecule comprising a suitable vector and an insert to be transcribed;

transcription of the DNA molecule utilizing elements required for transcription; and

purifying the transcribed mRNA product.

168. The method of Claim 167, wherein the elements required for transcription comprise a DNA-dependent RNA polymerase.

169. The method of Claim 167, wherein the elements required for transcription include wheat germ or reticulocyte extracts.

170. The method of Claim 167, further comprising the step of protecting the mRNA template.

171. The template formed by the method of Claim 170.

172. A method for forming a chimeric polynucleotide comprising:

preparing a single-stranded polynucleotide template;

modifying the single-stranded polynucleotide template such that the degree of chimeragenesis is increased;

contacting the modified template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

filling in gaps between hybridized oligonucleotides on the template; and

ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide.

173. The method of Claim 172 wherein, the population of oligonucleotides comprise a random population of oligonucleotides.

174. The method of Claim 172, wherein the modification is directed to filling in gaps.

175. The method of Claim 172, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

176. The method of Claim 172, further comprising a step of trimming flaps.

177. The method of Claim 172, wherein the modification comprises a chemical modification.

178. The method of Claim 176, wherein the template contains a plurality of uracil residues.

179. The method of Claim 177, wherein the modification creates abasic residues on the template.

180. The method of Claim 178, wherein the modification is a result of treatment with a uracil glycosylase.

181. The method of Claim 172, wherein the modification comprises a physical modification.

182. The method of Claim 180, wherein the modification is the hybridization of a pre-selected oligonucleotide to the template.

183. The method of Claim 181, wherein the pre-selected oligonucleotide is hybridized at a terminus of the template.

184. The method of Claim 182, wherein the pre-selected oligonucleotide is resistant to nucleotide extension.

185. The method of Claim 183, wherein the pre-selected oligonucleotide consists of a terminal dideoxy nucleoside.

186. The method of Claim 183, further comprising the steps of:

removing the first pre-selected oligonucleotide hybridized to the end of the template;

hybridizing a second pre-selected oligonucleotide to the end of the template, and

filling in the gap between the second oligonucleotide and an adjacently hybridized oligonucleotide.

187. The method of Claim 185, wherein the second pre-selected oligonucleotide is extended by a polymerase.

188. The method of Claim 183, further comprising the steps of:

hybridizing a second oligonucleotide immediately adjacent to the first oligonucleotide; and

filling in the gap between the second oligonucleotide and an adjacently hybridized oligonucleotide.

189. A method for forming a chimeric polynucleotide comprising:

preparing a single-stranded polynucleotide template containing a plurality of uracil residues;

contacting the template with a population of oligonucleotides, wherein at least two of the oligonucleotides hybridize to the template;

treating the template with an enzyme;

filling in gaps between hybridized oligonucleotides on the template; and

ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide.

190. The method of Claim 189, wherein gaps are filled in using a polymerase selected from a group consisting of T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase.

191. The method of Claim 189, further comprising a step of trimming flaps.

192. The method of Claim 189, wherein the enzyme is a uracil glycosylase.

193. The method of Claim 189, wherein the template contains between 0.01 and 100% uracil residues.

194. A method for forming a chimeric polynucleotide comprising:

preparing a single-stranded polynucleotide template;

hybridizing a first pre-selected oligonucleotide to the template;

contacting the template with a population of oligonucleotides under conditions such that at least two oligonucleotides hybridize to the template;

filling in gaps between hybridized oligonucleotides on the template; and

ligating adjacently hybridized oligonucleotides to form the chimeric polynucleotide.

195. The method of Claim 194, wherein the pre-selected oligonucleotide is resistant to nucleotide extension.

196. The method of Claim 195, wherein the pre-selected oligonucleotide consists of a terminal dideoxy nucleoside.

197. The method of Claim 195, wherein the pre-selected oligonucleotide contains at least one uracil.

198. The method of Claim 197, further comprising the steps of:

removing the first pre-selected oligonucleotide after the gap fill-in step;

hybridizing a second pre-selected oligonucleotide to a terminus of the template; and

filling in the gap between the second oligonucleotide and an adjacently hybridized oligonucleotide.

199. The method of Claim 198, wherein the second pre-selected oligonucleotide is extended by a polymerase.

200. The method of Claim 198, wherein gaps are filled in using a polymerase selected from a group consisting of T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase.

201. The method of Claim 198, further comprising a step of trimming flaps.

202. The method of Claim 195, further comprising the steps of:

hybridizing a second oligonucleotide immediately adjacent to the first oligonucleotide; and

filling in the gap between the second oligonucleotide and an adjacently hybridized oligonucleotide.

203. The method of Claim 202, wherein gaps are filled in using a polymerase selected from a group consisting of T4 DNA polymerase, *Taq* DNA polymerase or *Pfu* polymerase.

204. The method of Claim 202, further comprising a step of trimming flaps.

205. A chimeric polynucleotide prepared according to the method of Claim 172.

206. A method for preparing a modified single-stranded polynucleotide template suitable for use in forming a chimeric polynucleotide with an increased degree of chimeragenesis relative to a chimeric polynucleotide formed using an un-modified single-stranded polynucleotide template, comprising the steps of:

obtaining a single-stranded polynucleotide by a method selected from the group consisting of isolating a polynucleotide from a suitable nucleic acid source, synthetically manufacturing a polynucleotide, cleaving the polynucleotide from a larger polynucleotide, and amplifying a polynucleotide obtained by any of these methods; and

treating the single-stranded template such that the template is modified

an a manner that the step of filling in gaps is altered, thereby increasing the degree of chimeragenesis of the polynucleotide product obtained as a result of the present invention.

207. The method of Claim 206, wherein the modification is a chemical modification.

208. The method of Claim 207, wherein the chemical modification creates a-basic residues in the template.

209. The method of Claim 206, wherein the modification is a physical modification.

210. The method of Claim 209, wherein the modification is the hybridization of a pre-selected oligonucleotide to the template.

211. The method of Claim 210, wherein the pre-selected oligonucleotide is hybridized at one end of the template.

212. A method for forming a chimeric polynucleotide comprising:

contacting a single-stranded template comprising a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template; and

treating the hybridized oligonucleotides such that at least one contiguous chimeric polynucleotide is formed.

213. The method of Claim 212, wherein the phage vector is derived from bacteriophage M13.

214. The method of Claim 212, wherein the single-stranded template is isolated from a biological source.

215. The method of Claim 214, wherein the biological source is a bacteriophage grown in a bacterial strain defective in one or more dUTPases and uracil N-glycosylases.

216. The method of Claim 215, wherein the biological source is a bacteriophage grown in a bacterial strain containing a genome with a dut^{-}/ung^{-} genotype.

217. The method of Claim 212, wherein the single-stranded template contains at least one uracil residue.

218. The method of Claim 212, wherein the hybridized oligonucleotides are treated enzymatically.

219. The method of Claim 212, wherein the template and at least one parent polynucleotide used to form the population of oligonucleotides encodes a desulfurizing enzyme or a fragment thereof.

220. The method of Claim 212, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

221. The method of Claim 212, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

222. The method of Claim 212, wherein the single-stranded template comprises at least one region of random or partially random sequence.

223. The method of Claim 212, wherein the template or the population of oligonucleotides are generated from a chimeric polynucleotide produced by the method of Claim 212.

224. The method of Claim 212, further comprising at least one step selected from the group consisting of trimming flaps, filling in gaps between hybridized oligonucleotides and ligating hybridized oligonucleotides.

225. The method of Claim 224, wherein the step of trimming flaps precedes the step of filling in gaps between hybridized oligonucleotides.

226. The method of Claim 224, wherein the step of trimming flaps occurs concurrently with the step of filling in gaps between hybridized oligonucleotides.

227. The method of Claim 224, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

228. The method of Claim 224, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

229. A method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule are altered in comparison to at least one reference polynucleotide, comprising the steps of:

contacting a single-stranded template comprising a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

trimming flaps;

filling in gaps between hybridized oligonucleotides;

ligating hybridized oligonucleotides to form at least one chimeric polynucleotide;

selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide.

230. The method of Claim 229, wherein the phage vector is derived from bacteriophage M13.

231. The method of Claim 229, wherein the single-stranded template contains at least one uracil residue.

232. The method of Claim 229, wherein the hybridized oligonucleotides are treated enzymatically.

233. The method of Claim 229, wherein the template and at least one parent polynucleotide used to form the population of oligonucleotides encodes a desulfurizing enzyme or fragment thereof.

234. The method of Claim 229, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

235. The method of Claim 229, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

236. The method of Claim 229, wherein the single-stranded template comprises at least one region of random or partially random sequence.

237. The method of Claim 229, wherein the template or the population of oligonucleotides are generated from a chimeric polynucleotide produced by the method of Claim 229.

238. A method for generating at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule are altered in comparison to at least one reference polynucleotide, comprising the steps of:

contacting the single-stranded template comprising a phage vector and a polynucleotide insert with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

treating the hybridized oligonucleotides such that a heteroduplex polynucleotide consisting of a contiguous chimeric polynucleotide and the single-stranded template is formed; and

transforming a bacterial strain with the heteroduplex.

239. The method of Claim 238, wherein the phage vector is derived from bacteriophage M13.

240. The method of Claim 238, further comprising steps selected from the group consisting of trimming flaps, filling in gaps and ligating immediately adjacently hybridized oligonucleotides.

241. The method of Claim 238, further comprising the steps of:

selectively amplifying the chimeric polynucleotide over the single-stranded template; and

selecting chimeric polynucleotides with optimized properties when compared to the reference polynucleotide.

242. A method for the production of a library of chimeric polynucleotides comprising the steps of:

preparing a plurality of single-stranded templates comprising a phage vector and a polynucleotide insert;

contacting the single-stranded templates with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to more than one template; and

treating the oligonucleotides hybridized to each template such that more than one contiguous chimeric polynucleotide is formed, thereby generating the library of chimeric polynucleotides.

243. A library of chimeric polynucleotides produced by the method of Claim 242.

244. A method for producing a single-stranded polynucleotide template for use in forming a chimeric polynucleotide comprising the steps of:

inserting a polynucleotide into a phage vector;

transforming bacterial strain with the vector containing the insert;

inducing phage growth;

harvesting phage particles; and

isolating single-stranded DNA from the phage particles,

thereby resulting in the recovery of the single-stranded polynucleotide template.

245. The single-stranded polynucleotide produced according to the method of Claim 244.

246. A method of generating a population of single-stranded oligonucleotides for use in directed evolution, the method comprising:

obtaining a first double-stranded nucleic acid donor;

obtaining a second double-stranded nucleic acid, and incorporating uracil residues into the second double-stranded nucleic acid;

treating both double-stranded nucleic acids with lambda exonuclease to degrade the strand in each that contains a 5' phosphate, thereby obtaining

from each double-stranded nucleic acid a 5'strand without a phosphate group;

annealing the 5' strands to form heteroduplex nucleic acids;

treating the heteroduplex nucleic acids with an enzyme that cleaves mismatches to yield homoduplexes; and

treating the homoduplexes with an enzyme to degrade the strand containing the incorporated uracils; thereby generating a population of single-stranded oligonucleotides for use in directed evolution.

247. The method of Claim 246, wherein the enzyme that cleaves mismatches is a branch resolving enzyme.

248. The method of Claim 247, wherein the branch resolving enzyme is T4 endonuclease VII or T7 endonuclase I.

249. The method of Claim 246, wherein the enzyme used to degrade the strand containing the incorporated uracils is uracil DNA glycosylase.

250. The method of Claim 246, further comprising size fractionating the single-stranded oligonucleotides.

251. The method of Claim 246, wherein the first and second double-stranded nucleic acid are from a single source.

252. A method for forming a chemically modified single-stranded polynucleotide template for use in a method of directed evolution comprising:

preparing a double-stranded polynucleotide comprising the single-stranded polynucleotide template and a complementary polynucleotide strand;

denaturing the double-stranded polynucleotide;

adding a single-stranded oligonucleotide capable of annealing to the strand complementary to the single-stranded template; and

isolating the single-stranded polynucleotide template from its complementary strand and from the added oligonucleotide, thus yielding the purified single-stranded polynucleotide template.

253. The method of Claim 252, wherein the double-stranded polynucleotide comprising the single-stranded polynucleotide template and a complementary polynucleotide strand is a product of a nucleic acid amplification reaction.

254. The method of Claim 252, wherein a plurality of oligonucleotides capable of annealing to different sequences in the strand complementary to the single-stranded template are added after the step of denaturation.

255. The method of Claim 252, wherein the oligonucleotide capable of annealing to the strand complementary to the single-stranded template is comprised of RNA.

256. The method of Claim 252, wherein the olignucleotide capable of annealing to the strand complementary to the single-stranded template is preferably at least five nucleotides in length, but is shorter than the polynucleotide strand complementary to the single-stranded template.

257. The method of Claim 252, wherein the olignucleotide capable of annealing to the strand complementary to the single-stranded template is longer than the polynucleotide strand complementary to the single-stranded template

258. The method of Claim 252, wherein the single-stranded polynucleotide template is separated from its complement and added oligonucleotide by electrophoresis.

259. The method of Claim 252, wherein the single-stranded polynucleotide template is separated from its complement and added oligonucleotide by size exclusion chromatography.

260. The method of Claim 252, wherein the single-stranded polynucleotide template is separated from its complement and added oligonucleotide by ion-exchange chromatography.

261. The method of Claim 252, wherein the single-stranded polynucleotide template is separated from its complement and added oligonucleotide by affinity chromatography.

262. The method of Claim 252, wherein the single-stranded template comprises at least one region of random or partially random sequence.

263. A method for generating a chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide comprising the steps of:

preparing a double-stranded polynucleotide comprising a single-stranded polynucleotide template and a complementary polynucleotide strand,

denaturing the double-stranded polynucleotide;

adding a single-stranded oligonucleotide capable of annealing to the strand complementary to the single-stranded template;

isolating the single-stranded polynucleotide template from its complementary strand and from the added oligonucleotide, thus yielding the purified single-stranded polynucleotide template;

contacting the single-stranded template with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

trimming flaps;

filling in gaps;

ligating hybridized oligonucleotides to form at least one chimeric polynucleotide;

selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening the chimeric polynucleotide, wherein a characteristic is altered in comparison to the reference polynucleotide.

264. The method of Claim 263, further comprising selecting a chimeric polynucleotide for a particular characteristic.

265. The method of Claim 263, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

266. The method of Claim 263, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

267. The method of Claim 263, wherein the template contains a plurality of uracil residues.

268. The method of Claim 263, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* amplification or synthetic production.

269. The method of Claim 263, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

270. The method of Claim 263, wherein the single-stranded template comprises at least one region of random or partially random sequence.

271. A single-stranded template generated by the method of Claim 252.

272. A method for forming a chemically modified single-stranded polynucleotide template for use in a method of directed evolution comprising:

preparing a double-stranded polynucleotide comprising the chemically modified single-stranded polynucleotide template and a complementary polynucleotide strand;

denaturing the double-stranded polynucleotide; and

isolating the single-stranded polynucleotide template from its complementary strand, thus yielding the purified single-stranded polynucleotide template.

273. The method of Claim 272, wherein the chemical modification to the single-stranded template is a modification selected from the group consisting of biotinylation,

274. The method of Claim 272, wherein the double-stranded polynucleotide is prepared using an *in vitro* amplification reaction.

275. The method of Claim 272, wherein the isolation of the single-stranded template comprises the steps of:

immobilizing the single-stranded template;

separating the single-stranded template from its complementary strand; and

washing the complementary strand away from the immobilized single-stranded template, thus leaving the purified single-stranded template.

276. The method of Claim 275, wherein the single-stranded template is immobilized on a support selected from the group consisting of chromatography matrices, tube walls, slides and membranes.

277. A method for generating a chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide, comprising the steps of:

contacting the single-stranded template comprising a chemically modified single-stranded polynucleotide, which modification allows the single-stranded template to be immobilized on a solid matrix, and sequences derived from parent polynucleotides, with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening the chimeric polynucleotide, wherein a characteristic is altered in comparison to the reference polynucleotide.

278. The method of Claim 277, further comprising steps selected from the group consisting of trimming flaps, filling in gaps between hybridized oligonucleotides and ligating hybridized oligonucleotides.

279. The method of Claim 277, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* nucleic acid amplification or synthetic production.

280. The method of Claim 277, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

281. The method of Claim 277, wherein the single-stranded template comprises at least one region of random or partially random sequence.

282. The method of Claim 277, further comprising selecting a chimeric polynucleotide for a particular characteristic.

283. The method of Claim 278, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

284. The method of Claim 278, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

285. A chimeric polynucleotide prepared according to the method of Claim 277.

286. The template formed by the method of Claim 272.

287. A method for generating at least one chimeric polynucleotide, wherein at least one specific characteristic of the chimeric molecule is altered in comparison to at least one reference polynucleotide, comprising the steps of:

contacting a single-stranded template comprising sequences derived from at least one parent polynucleotide and at least one additional sequence whereby the single-stranded template is about 25% larger than a parent polynucleotide, with a population of oligonucleotides, under conditions such that at least two of the oligonucleotides hybridize to the template;

filling in gaps;

trimming flaps;

ligating the hybridized oligonucleotides, such that a chimeric polynucleotide is generated;

selectively amplifying the chimeric polynucleotide with respect to the single-stranded polynucleotide template; and

selecting or screening at least one chimeric polynucleotide, wherein a specified characteristic is altered in comparison to the reference polynucleotide.

288. The method of Claim 287, wherein the additional sequence is located on the 5' end relative to the parent-derived sequence.

289. The method of Claim 287, wherein the additional sequence is located on the 3' end relative to the parent-derived sequence.

290. The method of Claim 287, wherein the additional sequence is located within the parent-derived polynucleotide sequence.

291. The method of Claim 287, wherein a plurality of additional sequences are located at different positions in the template.

292. The method of Claim 287, wherein the single-stranded template comprises at least one region of random or partially random sequence.

293. The method of Claim 287, wherein the single-stranded template comprises a plurality of parent-derived sequences separated by linker sequences.

294. The method of Claim 287, wherein the selective amplification of the chimeric polynucleotide relative to the single-stranded template occurs by a method selected from the group consisting of electrophoresis, size exclusion chromatography, affinity chromatography, and ion-exchange chromatography.

295. The method of Claim 287, wherein the population of oligonucleotides includes oligonucleotides generated by *in vitro* nucleic acid amplification or synthetic production.

296. The method of Claim 287, wherein at least one oligonucleotide comprises a region of random or partially random sequence.

297. The method of Claim 287, further comprising steps selected from the group consisting of trimming flaps, filling in gaps between hybridized oligonucleotides and ligating hybridized oligonucleotides.

298. The method of Claim 296, wherein gaps are filled in using a polymerase selected from the group consisting of T4 DNA polymerase, *Taq* DNA polymerase and *Pfu* polymerase.

299. The method of Claim 296, wherein hybridized oligonucleotides are ligated using *Taq* DNA ligase or T4 DNA ligase.

300. The method of Claim 287, further comprising selecting a chimeric polynucleotide for a particular characteristic.

301. A chimeric polynucleotide prepared according to the method of Claim 287.

302. The template formed by the method of Claim 287.

303. The method of Claim 287, further comprising:

contacting the single-stranded template with an oligonucleotide comprising a plurality of uracils and sequences complementary to the 5' end of the single-stranded template after the polymerization step;

treating the polynucleotide with an uracil glycosylase; and

degrading polynucleotides comprising abasic sites.

304. A chimeric polynucleotide prepared according to the method of Claim 303.

305. A method for forming and selecting at least one chimeric polynucleotide, wherein one or more characteristics of the chimeric molecule is altered in comparison to a reference polynucleotide, comprising the steps of:

contacting a least one single-stranded polynucleotide template with a population of oligonucleotides under conditions wherein at least two oligonucleotides hybridize to the template;

treating hybridized oligonucleotides to form one chimeric polynucleotide hybridized to a template;

selectively amplifying the chimeric polynucleotide with respect to the templates; and

selecting or screening at least one chimeric polynucleotide, wherein the specified characteristic is altered in comparison to the reference polynucleotide.

306. The method of Claim 305, wherein the single-stranded template comprises a restriction endonuclease site at the 5' end.

307. The method of Claim 306, further comprising the step of treating the hybrid polynucleotide comprising the single-stranded template and the chimeric polynucleotide with a restriction endonuclease that cleaves double-stranded polynucleotides at the site in the 5' end of the template.

308. A method for forming a chimeric polynucleotide comprising:

preparing a random population of oligonucleotides from at least one nucleic acid with a preselected nucleotide sequence;

contacting a single-stranded template with the population of oligonucleotides under conditions such that at least two of the oligonucleotides hybridize to the template; and

ligating the hybridized oligonucleotides such that a chimeric polynucleotide is formed.

309. The method of Claim 308, further comprising a step of filling in gaps.

310. The method of Claim 309, wherein the oligonucletides comprise lengths in the range from about 10 nucleotides to about 200 nucleotides.

311. The method of Claim 310, wherein the hybridization conditions have been optimized to promote the annealing of short oligonucleotides.

312. The method of Claim 311, wherein the treatment utilizes a cryophilic enzyme.

313. A method for producing a single-stranded polynucleotide template for use in forming a chimeric polynucleotide comprising the steps of:

transcribing a double-stranded polynucleotide comprising a phage promoter to create an RNA transcript;

reverse transcribing the RNA transcript to create a hybrid polynucleotide; and

degrading the RNA strand of the hybrid polynucleotide,

thus resulting in a single-stranded polynucleotide template for use in forming a chimeric polynucleotide.

314. The single-stranded polynucleotide produced according to the method of Claim 313.

315. A method of forming a randomly fragmented population oligonucleotides for use in forming at least one chimeric polynucleotide comprising:

treating a double-stranded polynucleotide comprising a parent polynucleotide such that a plurality of modified bases are formed on either or both strands; and

treating the double-stranded polynucleotide that such stat single-strand nicks are created as a result of the treatment and the modified bases,

thus forming a randomly fragmented population of oligonucleotides for use in forming at least one chimeric polynucleotide.

316. The single-stranded polynucleotide produced according to the method of Claim 315.

317. A method of forming a population of single-stranded oligonucleotides for use in forming at least one chimeric polynucleotide comprising:

obtaining at least two double-stranded polynucleotides;

treating the double-stranded polynucleotides with an exonuclease, thereby obtaining from each double-stranded polynucleotide a strand containing a 3' phosphate;

annealing these strands to form heteroduplex nucleic acids;

treating the heteroduplex nucleic acids with an enzyme that cleaves mismatches to yield homoduplexes; and

treating the homoduplexes with an enzyme to degrade the strand containing the incorporated uracils,

thereby forming a population of single-stranded oligonucleotides for use in directed evolution.

318. The method of Claim 317, wherein the enzyme that cleaves mismatches is a branch resolving enzyme.

319. The method of Claim 318, wherein the branch resolving enzyme is T4 endonuclease VII or T7 endonuclase I.

320. The method of Claim 317, further comprising fractionation of the size single-stranded oligonucleotides based on size.

321. The method of Claim 317, wherein the first and second double-stranded nucleic acid are from a single source.

322. The population of single-stranded oligonucleotides produced according to the method of Claim 317.

323. A method of forming a population of oligonucleotides for use in forming at least one chimeric polynucleotide comprising:

obtaining a double-stranded nucleic acid polynucleotide;

obtaining a second double-stranded nucleic acid;

denaturing both double-stranded nucleic acids thereby obtaining single-stranded polynucleotides;

annealing these polynucleotides to form heteroduplex nucleic acids; and

treating the heteroduplex nucleic acids with an enzyme that cleaves mismatches to yield oligonucleotides,

thereby forming a population of oligonucleotides for use in forming at least one chimeric polynucleotide.

324. The population of single-stranded oligonucleotides produced according to the method of Claim 323.

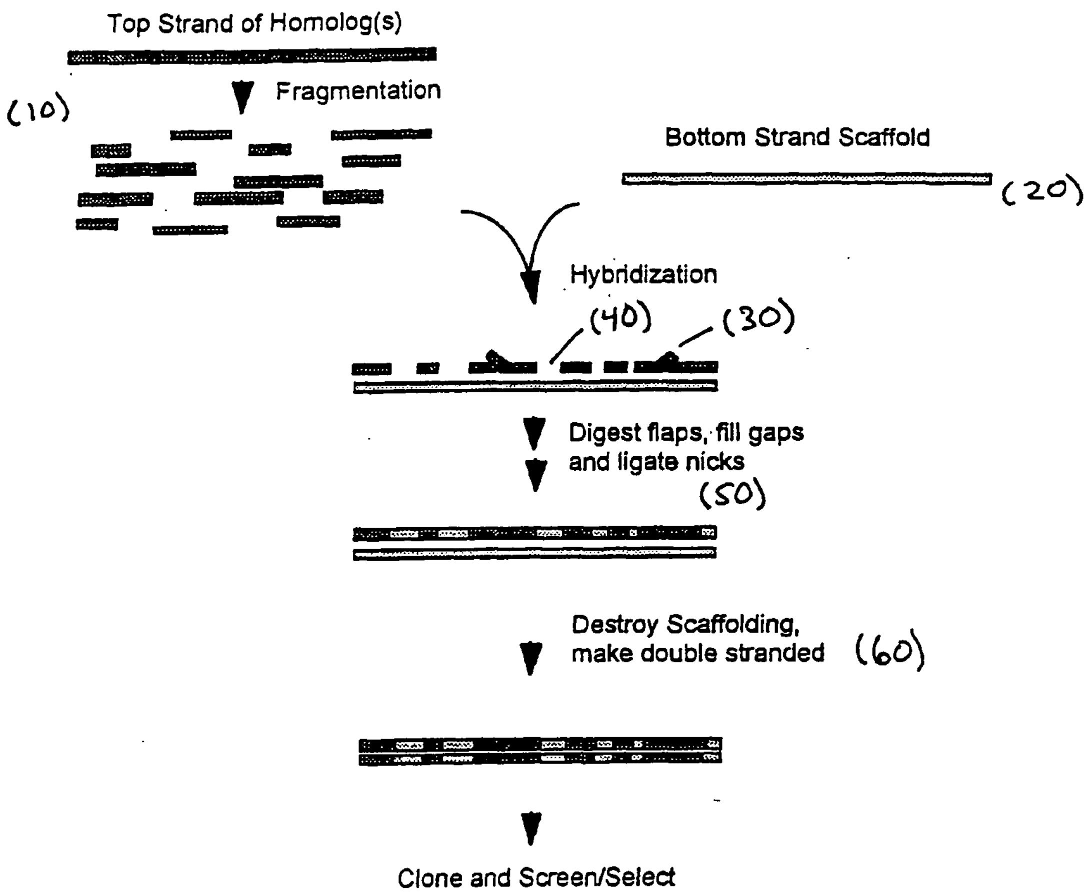

Top Strand of Homolog(s)
Fragmentation
(10)
Bottom Strand Scaffold
(20)
Hybridization
(40)
(30)
Digest flaps, fill gaps
and ligate nicks
(50)
Destroy Scaffolding,
make double stranded
(60)
Clone and Screen/Select

FIGURE 1.

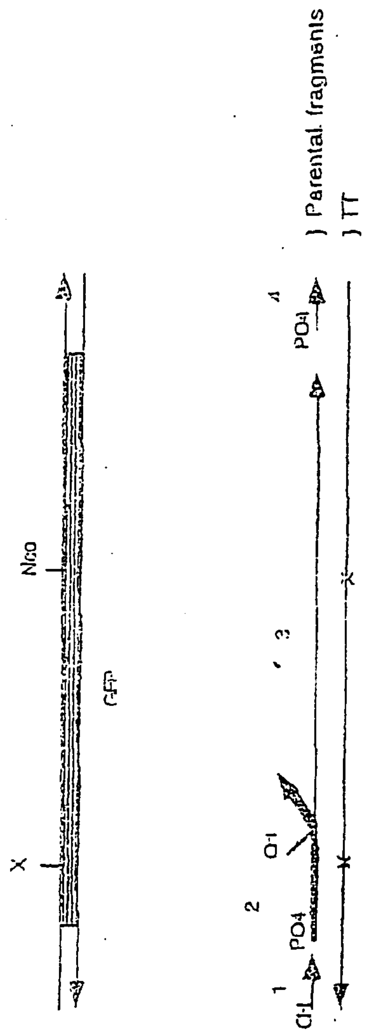


Figure 2

```
                                                          30                                                           60
ATG ACT CAA CAA CGA CAA ATG CAT CTG GCC GGT TTC TTC TCG GCC GGC AAT GTG ACT CAT
Met Thr Gln Gln Arg Gln Met His Leu Ala Gly Phe Phe Ser Ala Gly Asn Val Thr His

                                                          90                                                          120
GCA CAT GGG GCG TGG CGG CAC ACC GAC GCG TCG AAT GAC TTT CTG TCG GGG AAG TAC TAC
Ala His Gly Ala Trp Arg His Thr Asp Ala Ser Asn Asp Phe Leu Ser Gly Lys Tyr Tyr

                                                         150                                                          180
CAA CAC ATC GCC CGT ACT CTG GAG CGC GGC AAG TTC GAT CTG TTG TTT CTG CCT GAC GGG
Gln His Ile Ala Arg Thr Leu Glu Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Gly

                                                         210                                                          240
TTG GCC GTC GAG GAC AGC TAC GGG GAC AAC CTG GAC ACC GGT GTC GGC CTG GGC GGG CAG
Leu Ala Val Glu Asp Ser Tyr Gly Asp Asn Leu Asp Thr Gly Val Gly Leu Gly Gly Gln

                                                         270                                                          300
GGT GCA GTC GCC TTG GAG CCG GCC AGT GTG GTC GCA ACC ATG GCC GCG GTG ACC GAG CAC
Gly Ala Val Ala Leu Glu Pro Ala Ser Val Val Ala Thr Met Ala Ala Val Thr Glu His

                                                         330                                                          360
CTG GGT CTT GGG GCA ACC ATT TCG GCC ACC TAC TAT CCC CCG TAT CAC GTT GCT CGG GTG
Leu Gly Leu Gly Ala Thr Ile Ser Ala Thr Tyr Tyr Pro Pro Tyr His Val Ala Arg Val

                                                         390                                                          420
TTC GCG ACG CTC GAT CAG TTG TCA GGG GGT CGG GTG TCC TGG AAC GTC GTC ACC TCG CTC
Phe Ala Thr Leu Asp Gln Leu Ser Gly Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu
```

Figure 3A

```
                                450                                              480
AAC GAC GCT GAA GCG CGC AAC TTC GGC ATT AAT CAG CAT CTG GAA CAC GAC GCC CGC TAT
Asn Asp Ala Glu Ala Arg Asn Phe Gly Ile Asn Gln His Leu Glu His Asp Ala Arg Tyr

                                510                                              540
GAC CGC GCC GAT GAG TTC TTG GAA GCG GTC AAG AAA CTC TGG AAC AGC TGG GAC GAG GAC
Asp Arg Ala Asp Glu Phe Leu Glu Ala Val Lys Lys Leu Trp Asn Ser Trp Asp Glu Asp

                                570                                              600
GCC CTC GTG CTG GAC AAG GCG GCC GGC GTG TTC GCC GAT CCC GCG AAG GTG CAC TAC GTC
Ala Leu Val Leu Asp Lys Ala Ala Gly Val Phe Ala Asp Pro Ala Lys Val His Tyr Val

                                630                                              660
GAT CAC CAC GGG GAG TGG CTG AAT GTG CGC GGA CCT CTG CAG GTA CCG CGT TCA CCT CAG
Asp His His Gly Glu Trp Leu Asn Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln

                                690                                              720
GGT GAG CCG GTG ATC CTG CAG GCC GGC CTG TCG CCC CGG GGT CGG CGC TTC GCC GGG AAG
Gly Glu Pro Val Ile Leu Gln Ala Gly Leu Ser Pro Arg Gly Arg Arg Phe Ala Gly Lys

                                750                                              780
TGG GCC GAG GCC GTC TTC AGT CTT GCA CCC AAC CTC GAG GTG ATG CAG GCC ACC TAC CAG
Trp Ala Glu Ala Val Phe Ser Leu Ala Pro Asn Leu Glu Val Met Gln Ala Thr Tyr Gln

                                810                                              840
GGC ATC AAA GCC GAG GTC GAC GCT GCG GGG CGC GAT CCC GAT CAG ACG AAA ATC TTC ACC
Gly Ile Lys Ala Glu Val Asp Ala Ala Gly Arg Asp Pro Asp Gln Thr Lys Ile Phe Thr
```

Figure 3B

```
                                        870                                  900
GCC GTG ATG CCG GTA CTC GGC GAA AGC CAG GCG GTG GCA CAG GAA CGA CTG GAA TAT CTC
Ala Val Met Pro Val Leu Gly Glu Ser Gln Ala Val Ala Gln Glu Arg Leu Glu Tyr Leu

                                        930                                  960
AAC AGT CTG GTC CAT CCG GAA GTG GGA CTG TCG ACG CTA TCC AGT CAC ACC GGC ATC AAC
Asn Ser Leu Val His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Thr Gly Ile Asn

                                        990                                 1020
CTG GCG GCG TAC CCT CTC GAC ACT CCG ATC AAG GAC ATC CTG CGG GAT CTG CAG GAT CGG
Leu Ala Ala Tyr Pro Leu Asp Thr Pro Ile Lys Asp Ile Leu Arg Asp Leu Gln Asp Arg

                                       1050                                 1080
AAT GTC CCG ACG CAA CTG CAC ATG TTC GCC GCC GCA ACG CAC AGC GAA GAG CTC ACG CTG
Asn Val Pro Thr Gln Leu His Met Phe Ala Ala Ala Thr His Ser Glu Glu Leu Thr Leu

                                       1110                                 1140
GCG GAA ATG GGT CGG CGC TAT GGA ACC AAC GTG GGG TTC GTT CCT CAG TGG GCC GGT ACC
Ala Glu Met Gly Arg Arg Tyr Gly Thr Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr

                                       1170                                 1200
GGG GAG CAG ATC GCT GAC GAG CTG ATC CGC CAC TTC GAG GGC GGC GCC GCC GAT GGT TTC
Gly Glu Gln Ile Ala Asp Glu Leu Ile Arg His Phe Glu Gly Gly Ala Ala Asp Gly Phe

                                       1230                                 1260
ATC ATC TCT CCG GCC TTC CTG CCG GGC TCC TAC GAC GAG TTC GTC GAC CAG GTG GTT CCG
Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Asp Glu Phe Val Asp Gln Val Val Pro
```

Figure 3C

```
                                            1290                                       1320
GTT CTC CAG GAT CGC GGC TAC TTC CGC ACC GAG TAC CAG GGC AAC ACT CTG CGC GAC CAC
Val Leu Gln Asp Arg Gly Tyr Phe Arg Thr Glu Tyr Gln Gly Asn Thr Leu Arg Asp His

                                            1350
TTG GGT CTG CGC GTA CCA CAA CTG CAA GGA CAA CCT TCA TGA
Leu Gly Leu Arg Val Pro Gln Leu Gln Gly Gln Pro Ser  *
```

Figure 3D

```
                                                30                                                     60
ATG ACA AGC CGC GTC GAC CCC GCA AAC CCC GGT TCA GAA CTC GAT TCC GCC ATC CGC GAC
Met Thr Ser Arg Val Asp Pro Ala Asn Pro Gly Ser Glu Leu Asp Ser Ala Ile Arg Asp

                                                90                                                    120
ACA CTG ACC TAC AGC AAC TGC CCG GTA CCC AAC GCT CTG CTC ACG GCA TCG GAA TCG GGC
Thr Leu Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu Thr Ala Ser Glu Ser Gly

                                               150                                                    180
TTC CTC GAC GCC GCC GGC ATC GAA CTC GAC GTC CTC AGC GGC CAG CAG GGC ACG GTT CAT
Phe Leu Asp Ala Ala Gly Ile Glu Leu Asp Val Leu Ser Gly Gln Gln Gly Thr Val His

                                               210                                                    240
TTC ACC TAC GAC CAG CCT GCC TAC ACC CGT TTT GGG GGT GAG ATC CCG CCA CTG CTC AGC
Phe Thr Tyr Asp Gln Pro Ala Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Leu Ser

                                               270                                                    300
GAG GGG TTG CGG GCA CCT GGG CGC ACG CGT CTA CTC GGC ATC ACC CCG CTC TTG GGG CGC
Glu Gly Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Ile Thr Pro Leu Leu Gly Arg

                                               330                                                    360
CAG GGC TTC TTT GTC CGC GAC GAC AGC CCG ATC ACA GCG GCC GCC GAC CTT GCC GGA CGT
Gln Gly Phe Phe Val Arg Asp Asp Ser Pro Ile Thr Ala Ala Ala Asp Leu Ala Gly Arg

                                               390                                                    420
CGA ATC GGC GTC TCG GCC TCG GCA ATT CGC ATC CTG CGC GGC CAG CTG GGC GAC TAC CTC
Arg Ile Gly Val Ser Ala Ser Ala Ile Arg Ile Leu Arg Gly Gln Leu Gly Asp Tyr Leu
```

Figure 4A

```
                                        450                                    480
GAG TTG GAT CCC TGG CGG CAA ACG CTG GTA GCG CTG GGC TCG TGG GAG GCG CGC GCC TTG
Glu Leu Asp Pro Trp Arg Gln Thr Leu Val Ala Leu Gly Ser Trp Glu Ala Arg Ala Leu

                                        510                                    540
TTG CAC ACC CTT GAG CAC GGT GAA CTG GGT GTG GAC GAC GTC GAG CTG GTG CCG ATC AGC
Leu His Thr Leu Glu His Gly Glu Leu Gly Val Asp Asp Val Glu Leu Val Pro Ile Ser

                                        570                                    600
AGT CCT GGT GTC GAT GTT CCC GCT GAG CAG CTC GAA GAA TCG GCG ACC GTC AAG GGT GCG
Ser Pro Gly Val Asp Val Pro Ala Glu Gln Leu Glu Glu Ser Ala Thr Val Lys Gly Ala

                                        630                                    660
GAC CTC TTT CCC GAT GTC GCC CGC GGT CAG GCC GCG GTG TTG GCC AGC GGA GAC GTT GAC
Asp Leu Phe Pro Asp Val Ala Arg Gly Gln Ala Ala Val Leu Ala Ser Gly Asp Val Asp

                                        690                                    720
GCC CTG TAC AGT TGG CTG CCC TGG GCC GGG GAG TTG CAA GCC ACC GGG GCC CGC CCA GTG
Ala Leu Tyr Ser Trp Leu Pro Trp Ala Gly Glu Leu Gln Ala Thr Gly Ala Arg Pro Val

                                        750                                    780
GTG GAT CTC GGC CTC GAT GAG CGC AAT GCC TAC GCC AGT GTG TGG ACG GTC AGC AGC GGG
Val Asp Leu Gly Leu Asp Glu Arg Asn Ala Tyr Ala Ser Val Trp Thr Val Ser Ser Gly

                                        810                                    840
CTG GTT CGC CAG CGA CCT GGC CTT GTT CAA CGA CTG GTC GAC GCG GCC GTC GAC GCC GGG
Leu Val Arg Gln Arg Pro Gly Leu Val Gln Arg Leu Val Asp Ala Ala Val Asp Ala Gly
```

Figure 4B

```
                                    870                             900
CTG TGG GCA CGC GAT CAT TCC GAC GCG GTG ACC AGC CTG CAC GCC GCG AAC CTG GGC GTA
Leu Trp Ala Arg Asp His Ser Asp Ala Val Thr Ser Leu His Ala Ala Asn Leu Gly Val

                                    930                             960
TCG ACC GGA GCA GTA GGC CAG GGC TTC GGC GCC GAC TTC CAG CAG CGT CTG GTT CCA CGC
Ser Thr Gly Ala Val Gly Gln Gly Phe Gly Ala Asp Phe Gln Gln Arg Leu Val Pro Arg

                                    990                            1020
CTG GAT CAC GAC GCC CTC GCC CTC CTG GAG CGC ACA CAG CAA TTC CTG CTC ACC AAC AAC
Leu Asp His Asp Ala Leu Ala Leu Leu Glu Arg Thr Gln Gln Phe Leu Leu Thr Asn Asn

                                   1050                            1080
TTG CTG CAG GAA CCC GTC GCC CTC GAT CAG TGG GCG GCT CCG GAA TTT CTG AAC AAC AGC
Leu Leu Gln Glu Pro Val Ala Leu Asp Gln Trp Ala Ala Pro Glu Phe Leu Asn Asn Ser

CTC AAT CGC CAC CGA TAG
Leu Asn Arg His Arg  *
```

Figure 4C

```
                                                30                                                      60
ATG ACA CTG TCA CCT GAA AAG CAG CAC GTT CGA CCA CGC GAC GCC GCC GAC AAC GAT CCC
Met Thr Leu Ser Pro Glu Lys Gln His Val Arg Pro Arg Asp Ala Ala Asp Asn Asp Pro

                                                90                                                     120
GTC GCG GTT GCC CGT GGG CTA GCC GAA AAG TGG CGA GCC ACC GCC GTC GAG CGT GAT CGC
Val Ala Val Ala Arg Gly Leu Ala Glu Lys Trp Arg Ala Thr Ala Val Glu Arg Asp Arg

                                               150                                                     180
GCC GGG GGT TCG GCA ACA GCC GAG CGC GAA GAC CTG CGC GCG AGC GCG CTG CTG TCG CTC
Ala Gly Gly Ser Ala Thr Ala Glu Arg Glu Asp Leu Arg Ala Ser Ala Leu Leu Ser Leu

                                               210                                                     240
CTC GTC CCG CGC GAA TAC GGC GGC TGG GGC GCA GAC TGG CCC ACC GCC ATC GAG GTC GTC
Leu Val Pro Arg Glu Tyr Gly Gly Trp Gly Ala Asp Trp Pro Thr Ala Ile Glu Val Val

                                               270                                                     300
CGC GAA ATC GCG GCA GCC GAT GGA TCT TTG GGA CAC CTG TTC GGA TAC CAC CTC ACC AAC
Arg Glu Ile Ala Ala Ala Asp Gly Ser Leu Gly His Leu Phe Gly Tyr His Leu Thr Asn

                                               330                                                     360
GCC CCG ATG ATC GAA CTG ATC GGC TCG CAG GAA CAA GAA GAA CAC CTG TAC ACC CAG ATC
Ala Pro Met Ile Glu Leu Ile Gly Ser Gln Glu Gln Glu Glu His Leu Tyr Thr Gln Ile

                                               390                                                     420
GCG CAG AAC AAC TGG TGG ACC GGA AAT GCC TCC AGC GAG AAC AAC AGC CAC GTG CTG GAC
Ala Gln Asn Asn Trp Trp Thr Gly Asn Ala Ser Ser Glu Asn Asn Ser His Val Leu Asp
```

Figure 5A

```
                                            450                                        480
TGG AAG GTC AGC GCC ACC CCG ACC GAA GAC GGC GGC TAC GTG CTC AAT GGC ACG AAG CAC
Trp Lys Val Ser Ala Thr Pro Thr Glu Asp Gly Gly Tyr Val Leu Asn Gly Thr Lys His

                                            510                                        540
TTC TGC AGC GGC GCC AAG GGG TCG GAC CTG CTG TTC GTG TTC GGC GTC GTC CAG GAT GAT
Phe Cys Ser Gly Ala Lys Gly Ser Asp Leu Leu Phe Val Phe Gly Val Val Gln Asp Asp

                                            570                                        600
TCT CCG CAG CAG GGT GCG ATC ATT GCT GCC GCT ATC CCG ACA TCG CGG GCT GGC GTT ACG
Ser Pro Gln Gln Gly Ala Ile Ile Ala Ala Ala Ile Pro Thr Ser Arg Ala Gly Val Thr

                                            630                                        660
CCC AAC GAC GAC TGG GCC GCC ATC GGC ATG CGG CAG ACC GAC AGC GGT TCC ACG GAC TTC
Pro Asn Asp Asp Trp Ala Ala Ile Gly Met Arg Gln Thr Asp Ser Gly Ser Thr Asp Phe

                                            690                                        720
CAC AAC GTC AAG GTC GAG CCT GAC GAA GTG CTG GGC GCG CCC AAC GCC TTC GTT CTC GCC
His Asn Val Lys Val Glu Pro Asp Glu Val Leu Gly Ala Pro Asn Ala Phe Val Leu Ala

                                            750                                        780
TTC ATA CAA TCC GAG CGC GGC AGC CTC TTC GCG CCC ATA GCG CAA TTG ATC TTC GCC AAC
Phe Ile Gln Ser Glu Arg Gly Ser Leu Phe Ala Pro Ile Ala Gln Leu Ile Phe Ala Asn

                                            810                                        840
GTC TAT CTG GGG ATC GCG CAC GGC GCA CTC GAT GCC GCC AGG GAG TAC ACC CGT ACC CAG
Val Tyr Leu Gly Ile Ala His Gly Ala Leu Asp Ala Ala Arg Glu Tyr Thr Arg Thr Gln
```

Figure 5B

```
                                        870                                   900
GCG AGG CCC TGG ACA CCG GCC GGT ATT CAA CAG GCA ACC GAG GAT CCC TAC ACC ATC CGC
Ala Arg Pro Trp Thr Pro Ala Gly Ile Gln Gln Ala Thr Glu Asp Pro Tyr Thr Ile Arg

                                        930                                   960
TCC TAC GGT GAG TTC ACC ATC GCA TTG CAG GGA GCT GAC GCC GCC GCC CGT GAA GCG GCC
Ser Tyr Gly Glu Phe Thr Ile Ala Leu Gln Gly Ala Asp Ala Ala Ala Arg Glu Ala Ala

                                        990                                  1020
CAC CTG CTG CAG ACG GTG TGG GAC AAG GGC GAC GCG CTC ACC CCC GAG GAC CGC GGC GAA
His Leu Leu Gln Thr Val Trp Asp Lys Gly Asp Ala Leu Thr Pro Glu Asp Arg Gly Glu

                                       1050                                  1080
CTG ATG GTG AAG GTC TCG GGA GTC AAA GCG TTG GCC ACC AAC GCC GCC CTC AAC ATC AGC
Leu Met Val Lys Val Ser Gly Val Lys Ala Leu Ala Thr Asn Ala Ala Leu Asn Ile Ser

                                       1110                                  1140
AGC GGC GTC TTC GAG GTG ATC GGC GCG CGC GGA ACA CAT CCC AGG TAC GGT TTC GAC CGC
Ser Gly Val Phe Glu Val Ile Gly Ala Arg Gly Thr His Pro Arg Tyr Gly Phe Asp Arg

                                       1170                                  1200
TTC TGG CGC AAC GTG CGC ACC CAC TCC CTG CAC GAC CCG GTG TCC TAC AAG ATC GCC GAC
Phe Trp Arg Asn Val Arg Thr His Ser Leu His Asp Pro Val Ser Tyr Lys Ile Ala Asp

                                       1230
GTC GGC AAG CAC ACC TTG AAC GGT CAA TAC CCG ATT CCC GGC TTC ACC TCC TGA
Val Gly Lys His Thr Leu Asn Gly Gln Tyr Pro Ile Pro Gly Phe Thr Ser  *
```

Figure 5C

```
                                             30                                                60
ATG ACC GAT CCA CGT CAG CTG CAC CTG GCC GGA TTC TTC TGT GCC GGC AAC GTC ACG CAC
Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly Asn Val Thr His

                                             90                                               120
GCC CAC GGA GCG TGG CGC CAC GCC GAC GAC TCC AAC GGC TTC CTC ACC AAG GAG TAC TAC
Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn Gly Phe Leu Thr Lys Glu Tyr Tyr

                                            150                                               180
CAG CAG ATT GCC CGC ACG CTC GAG CGC GGC AAG TTC GAC CTG CTG TTC CTT CCC GAC GCG
Gln Gln Ile Ala Arg Thr Leu Glu Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala

                                            210                                               240
CTC GCC GTG TGG GAC AGC TAC GGC GAC AAT CTG GAG ACC GGT CTG CGG TAT GGC GGG CAA
Leu Ala Val Trp Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gly Gln

                                            270                                               300
GGC GCG GTG ATG CTG GAG CCC GGC GTA GTT ATC GCC GCG ATG GCC TCG GTG ACC GAA CAT
Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser Val Thr Glu His

                                            330                                               360
CTG GGG CTG GGC GCC ACC ATT TCC ACC ACC TAC TAC CCG CCC TAC CAT GTA GCC CGG GTC
Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr Pro Pro Tyr His Val Ala Arg Val

                                            390                                               420
GTC GCT TCG CTG GAC CAG CTG TCC TCC GGG CGA GTG TCG TGG AAC GTG GTC ACC TCG CTC
Val Ala Ser Leu Asp Gln Leu Ser Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu
```

Figure 6A

```
                                             450                                                    490
AGC AAT GCA GAG GCG CGC AAC TTC GGC TTC CAT GAA CAT CTC GAC CAC GAT GCC CGC TAC
Ser Asn Ala Glu Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr

                                             510                                                    540
GAT CGC GCC GAT GAA TTC CTC GAG GTC GTG CGC AAG CTC TGG AAC AGC TGG GAT CGC GAT
Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser Trp Asp Arg Asp

                                             570                                                    600
GCG CTG ACA CTC GAC AAG GCA ACC GGC CAG TTC GCC GAT CCG GCT AAG GTG CGC TAC ATC
Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala Asp Pro Ala Lys Val Arg Tyr Ile

                                             630                                                    660
GAC CAC CGC GGC GAA TGG CTC AAC GTA CGC GGG CCG CTT CAG GTG CCG CGC TCC CCC CAG
Asp His Arg Gly Glu Trp Leu Asn Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln

                                             690                                                    720
GGC GAG CCT GTC ATT CTG CAG GCC GGG CTT TCG GCG CGG GGC AAG CGC TTC GCC GGG CGC
Gly Glu Pro Val Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg

                                             750                                                    780
TGG GCG GAC GCG GTG TTC ACG ATT TCG CCC AAT CTG GAC ATC ATG CAG GCC ACG TAC CGC
Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln Ala Thr Tyr Arg

                                             810                                                    840
GAC ATA AAG GCG CAG GTC GAG GCC GCC GGA CGC GAT CCC GAG CAG GTC AAG GTG TTT GCC
Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp Pro Glu Gln Val Lys Val Phe Ala
```

Figure 6B

```
                                                      870                                                 900
GCG GTG ATG CCG ATC CTC GGC GAG ACC GAG GCG ATC GCC AGG CAG CGT CTC GAA TAC ATA
Ala Val Met Pro Ile Leu Gly Glu Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile

                                                      930                                                 960
AAT TCG CTG GTG CAT CCC GAA GTC GGG CTT TCT ACG TTG TCC AGC CAT GTC GGG GTC AAC
Asn Ser Leu Val His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn

                                                      990                                                1020
CTT GCC GAC TAT TCG CTC GAT ACC CCG CTG ACC GAG GTC CTG GGC GAT CTC GCC CAG CGC
Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp Leu Ala Gln Arg

                                                     1050                                                1080
AAC GTG CCC ACC CAA CTG GGC ATG TTC GCC AGG ATG TTG CAG GCC GAG ACG CTG ACC GTG
Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met Leu Gln Ala Glu Thr Leu Thr Val

                                                     1110                                                1140
GGA GAA ATG GGC CGG CGT TAT GGC GCC AAC GTG GGC TTC GTC CCG CAG TGG GCG GGA ACC
Gly Glu Met Gly Arg Arg Tyr Gly Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr

                                                     1170                                                1200
CGC GAG CAG ATC GCG GAC CTG ATC GAG ATC CAT TTC AAG GCC GGC GGC GCC GAT GGC TTC
Arg Glu Gln Ile Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe

                                                     1230                                                1260
ATC ATC TCG CCG GCG TTC CTG CCC GGA TCT TAC GAG GAA TTC GTC GAT CAG GTG GTG CCC
Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp Gln Val Val Pro
```

Figure 6C

```
                                                     1290                                            1320
ATC CTG CAG CAC CGC GGA CTG TTC CGC ACT GAT TAC GAA GGC CGC ACC CTG CGC AGC CAT
Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr Glu Gly Arg Thr Leu Arg Ser His

                                                     1350
CTG GGA CTG CGT GAA CCC GCA TAC CTG GGA GAG TAC GCA TGA
Leu Gly Leu Arg Glu Pro Ala Tyr Leu Gly Glu Tyr Ala  *
```

Figure 6D

```
                                             30                                                60
ATG ACG ACA GAC ATC CAC CCG GCG AGC GCC GCA TCG TCG CCG GCG GCG CGC GCG ACG ATC
Met Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala Arg Ala Thr Ile

                                             90                                               120
ACC TAC AGC AAC TGC CCC GTG CCT AAT GCC CTG CTC GCC GCG CTC GGC TCA GGT ATT CTG
Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu Ala Ala Leu Gly Ser Gly Ile Leu

                                            150                                               180
GAC AGT GCC GGG ATC ACA CTT GCC CTG CTG ACC GGA AAG CAG GGC GAG GTG CAC TTC ACC
Asp Ser Ala Gly Ile Thr Leu Ala Leu Leu Thr Gly Lys Gln Gly Glu Val His Phe Thr

                                            210                                               240
TAC GAC CGA GAT GAC TAC ACC CGC TTC GGC GGC GAG ATT CCG CCG CTG GTC AGC GAG GGA
Tyr Asp Arg Asp Asp Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Val Ser Glu Gly

                                            270                                               300
CTG CGT GCG CCG GGG CGG ACC CGC CTG CTG GGA CTG ACG CCG GTG CTG GGC CGC TGG GGC
Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Leu Thr Pro Val Leu Gly Arg Trp Gly

                                            330                                               360
TAC TTC GTC CGG GGC GAC AGC GCG ATC CGC ACC CCG GCC GAT CTT GCC GGC CGC CGC GTC
Tyr Phe Val Arg Gly Asp Ser Ala Ile Arg Thr Pro Ala Asp Leu Ala Gly Arg Arg Val

                                            390                                               420
GGA GTA TCC GAT TCG GCC AGG AGG ATA TTG ACC GGA AGG CTG GGC GAC TAC CGC GAA CTT
Gly Val Ser Asp Ser Ala Arg Arg Ile Leu Thr Gly Arg Leu Gly Asp Tyr Arg Glu Leu
```

Figure 7A

```
                                                450                                        490
GAT CCC TGG CGG CAG ACC CTG GTC GCG CTG GGG ACA TGG GAG GCG CGT GCC TTG CTG AGC
Asp Pro Trp Arg Gln Thr Leu Val Ala Leu Gly Thr Trp Glu Ala Arg Ala Leu Leu Ser

                                                510                                        540
ACG CTC GAG ACG GCG GGG CTT GGC GTC GGC GAC GTC GAG CTG ACG CGC ATC GAG AAC CCG
Thr Leu Glu Thr Ala Gly Leu Gly Val Gly Asp Val Glu Leu Thr Arg Ile Glu Asn Pro

                                                570                                        600
TTC GTC GAC GTG CCG ACC GAA CGA CTG CAT GCC GCC GGC TCG CTC AAA GGA ACC GAC CTG
Phe Val Asp Val Pro Thr Glu Arg Leu His Ala Ala Gly Ser Leu Lys Gly Thr Asp Leu

                                                630                                        660
TTC CCC GAC GTG ACC AGC CAG CAG GCC GCA GTC CTT GAG GAT GAG CGC GCC GAC GCC CTG
Phe Pro Asp Val Thr Ser Gln Gln Ala Ala Val Leu Glu Asp Glu Arg Ala Asp Ala Leu

                                                690                                        720
TTC GCG TGG CTT CCC TGC GCG GCC GAG CTC GAG ACC CGC ATC GGT GCA CGG CCG GTC CTA
Phe Ala Trp Leu Pro Trp Ala Ala Glu Leu Glu Thr Arg Ile Gly Ala Arg Pro Val Leu

                                                750                                        780
GAC CTC AGC GCA GAC GAC CGC AAT GCC TAT GCG AGC ACC TGG ACG GTG AGC GCC GAG CTG
Asp Leu Ser Ala Asp Asp Arg Asn Ala Tyr Ala Ser Thr Trp Thr Val Ser Ala Glu Leu

                                                810                                        840
GTG GAC CGG CAG CCC GAA CTG GTG CAG CGG CTC GTC GAT GCC GTG GTG GAT GCA GGG CGG
Val Asp Arg Gln Pro Glu Leu Val Gln Arg Leu Val Asp Ala Val Val Asp Ala Gly Arg
```

Figure 7B

```
                                            870                                    900
TGG GCC GAG GCC AAT GGC GAT GTC GTC TCC CGC CTG CAC GCC GAT AAC CTC GGT GTC AGT
Trp Ala Glu Ala Asn Gly Asp Val Val Ser Arg Leu His Ala Asp Asn Leu Gly Val Ser

                                            930                                    960
CCC GAA AGC GTC CGC CAG GGA TTC GGA GCC GAT TTT CAC CGC CGC CTG ACG CCG CGG CTC
Pro Glu Ser Val Arg Gln Gly Phe Gly Ala Asp Phe His Arg Arg Leu Thr Pro Arg Leu

                                            990                                   1020
GAC AGC GAT GCT ATC GCC ATC CTG GAG CGT ACT CAG CGG TTC CTG AAG GAT GCG AAC CTG
Asp Ser Asp Ala Ile Ala Ile Leu Glu Arg Thr Gln Arg Phe Leu Lys Asp Ala Asn Leu

                                           1050                                   1080
ATC GAT CGG TCG TTG GCG CTC GAT CGG TGG GCT GCA CCT GAA TTC CTC GAA CAA AGT CTC
Ile Asp Arg Ser Leu Ala Leu Asp Arg Trp Ala Ala Pro Glu Phe Leu Glu Gln Ser Leu

                                           1110
TCA CGC CAG GTC GAA GGG CAG ATA GCA TGA
 Ser Arg Gln Val Glu Gly Gln Ile Ala
```

Figure 7C

```
                                             30                                                  60
ATG AAC GAA CTC GTC AAA GAT CTC GGC CTC AAT CGA TCC GAT CCG ATC GGC GCT GTG CGG
Met Asn Glu Leu Val Lys Asp Leu Gly Leu Asn Arg Ser Asp Pro Ile Gly Ala Val Arg

                                             90                                                 120
CGA CTG GCC GCG CAG TGG GGG GCC ACC GCT GTT GAT CGG GAC CGG GCC GGC GGA TCG GCA
Arg Leu Ala Ala Gln Trp Gly Ala Thr Ala Val Asp Arg Asp Arg Ala Gly Gly Ser Ala

                                            150                                                 180
ACC GCC GAA CTC GAT CAA CTG CGC GGC AGC GGC CTG CTC TCG CTG TCC ATT CCC GCC GCA
Thr Ala Glu Leu Asp Gln Leu Arg Gly Ser Gly Leu Leu Ser Leu Ser Ile Pro Ala Ala

                                            210                                                 240
TAT GGC GGC TGG GGC GCC GAC TGG CCA ACG ACT CTG GAA GTT ATC CGC GAA GTC GCA ACG
Tyr Gly Gly Trp Gly Ala Asp Trp Pro Thr Thr Leu Glu Val Ile Arg Glu Val Ala Thr

                                            270                                                 300
GTG GAC GGA TCG CTG GCG CAT CTA TTC GGC TAC CAC CTC GGC TGC GTA CCG ATG ATC GAG
Val Asp Gly Ser Leu Ala His Leu Phe Gly Tyr His Leu Gly Cys Val Pro Met Ile Glu

                                            330                                                 360
CTG TTC GGC TCG GCG CCA CAA AAG GAA CGG CTG TAC CGC CAG ATC GCA AGC CAT GAT TGG
Leu Phe Gly Ser Ala Pro Gln Lys Glu Arg Leu Tyr Arg Gln Ile Ala Ser His Asp Trp

                                            390                                                 420
CGG GTC GGG AAT GCG TCG AGC GAA AAC AAC AGC CAC GTG CTC GAG TGG AAG CTT GCC GCC
Arg Val Gly Asn Ala Ser Ser Glu Asn Asn Ser His Val Leu Glu Trp Lys Leu Ala Ala
```

Figure 8A

```
                                             450                                                  480
ACC GCC GTC GAT GAT GGC GGG TTC GTC CTC AAC GGC GCG AAG CAC TTC TGC AGC GGC GCC
Thr Ala Val Asp Asp Gly Gly Phe Val Leu Asn Gly Ala Lys His Phe Cys Ser Gly Ala

                                             510                                                  540
AAA AGC TCC GAC CTG CTC ATC GTG TTC GGC GTG ATC CAG GAC GAA TCC CCC CTG CGC GGC
Lys Ser Ser Asp Leu Leu Ile Val Phe Gly Val Ile Gln Asp Glu Ser Pro Leu Arg Gly

                                             570                                                  600
GCG ATC ATC ACC GCG GTC ATT CCC ACC GAC CGG GCC GGT GTT CAG ATC AAT GAC GAC TGG
Ala Ile Ile Thr Ala Val Ile Pro Thr Asp Arg Ala Gly Val Gln Ile Asn Asp Asp Trp

                                             630                                                  660
CGC GCA ATC GGG ATG CGC CAG ACC GAC AGC GGC AGC GCC GAA TTT CGC GAC GTC CGA GTC
Arg Ala Ile Gly Met Arg Gln Thr Asp Ser Gly Ser Ala Glu Phe Arg Asp Val Arg Val

                                             690                                                  720
TAC CCA GAC GAG ATC TTG GGG GCA CCA AAC TCA GTC GTT GAG GCG TTC GTG ACA AGC AAC
Tyr Pro Asp Glu Ile Leu Gly Ala Pro Asn Ser Val Val Glu Ala Phe Val Thr Ser Asn

                                             750                                                  780
CGC GGC AGC CTG TGG ACG CCG GCG ATT CAG TCG ATC TTC TCG AAC GTT TAT CTG GGG CTC
Arg Gly Ser Leu Trp Thr Pro Ala Ile Gln Ser Ile Phe Ser Asn Val Tyr Leu Gly Leu

                                             810                                                  840
GCG CGT GGC GCG CTC GAG GCG GCA GCG GAT TAC ACC CGG ACC CAG AGC CGC CCC TGG ACA
Ala Arg Gly Ala Leu Glu Ala Ala Ala Asp Tyr Thr Arg Thr Gln Ser Arg Pro Trp Thr
```

Figure 8B

```
                                             870                                                   900
CCC GCC GGC GTG GCG AAG GCG ACA GAG GAT CCC CAC ATC ATC GCC ACC TAC GGT GAA CTG
Pro Ala Gly Val Ala Lys Ala Thr Glu Asp Pro His Ile Ile Ala Thr Tyr Gly Glu Leu

                                             930                                                   960
GCG ATC GCG CTC CAG GGC GCC GAG GCG GCC GCG CGC GAG GTC GCG GCC CTG TTG CAA CAG
Ala Ile Ala Leu Gln Gly Ala Glu Ala Ala Ala Arg Glu Val Ala Ala Leu Leu Gln Gln

                                             990                                                  1020
GCG TGG GAC AAG GGC GAT GCG GTG ACG CCC GAA GAG CGC GGC CAG CTG ATG GTG AAG GTT
Ala Trp Asp Lys Gly Asp Ala Val Thr Pro Glu Glu Arg Gly Gln Leu Met Val Lys Val

                                            1050                                                  1080
TCG GGT GTG AAG GCC CTC TCG ACG AAG GCC GCC CTC GAC ATC ACC AGC CGT ATT TTC GAG
Ser Gly Val Lys Ala Leu Ser Thr Lys Ala Ala Leu Asp Ile Thr Ser Arg Ile Phe Glu

                                            1110                                                  1140
ACA ACG GGC TCG CGA TCG ACG CAT CCC AGA TAC GGA TTC GAT CGG TTC TGG CGT AAC ATC
Thr Thr Gly Ser Arg Ser Thr His Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn Ile

                                            1170                                                  1200
CGG ACT CAT ACG CTG CAC GAT CCG GTA TCG TAT AAA ATC GTC GAT GTG GGG AAC TAC ACG
Arg Thr His Thr Leu His Asp Pro Val Ser Tyr Lys Ile Val Asp Val Gly Asn Tyr Thr

                                            1230
CTC AAC GGG ACA TTC CCG GTT CCC GGA TTT ACG TCA
Leu Asn Gly Thr Phe Pro Val Pro Gly Phe Thr Ser
```

Figure 8C

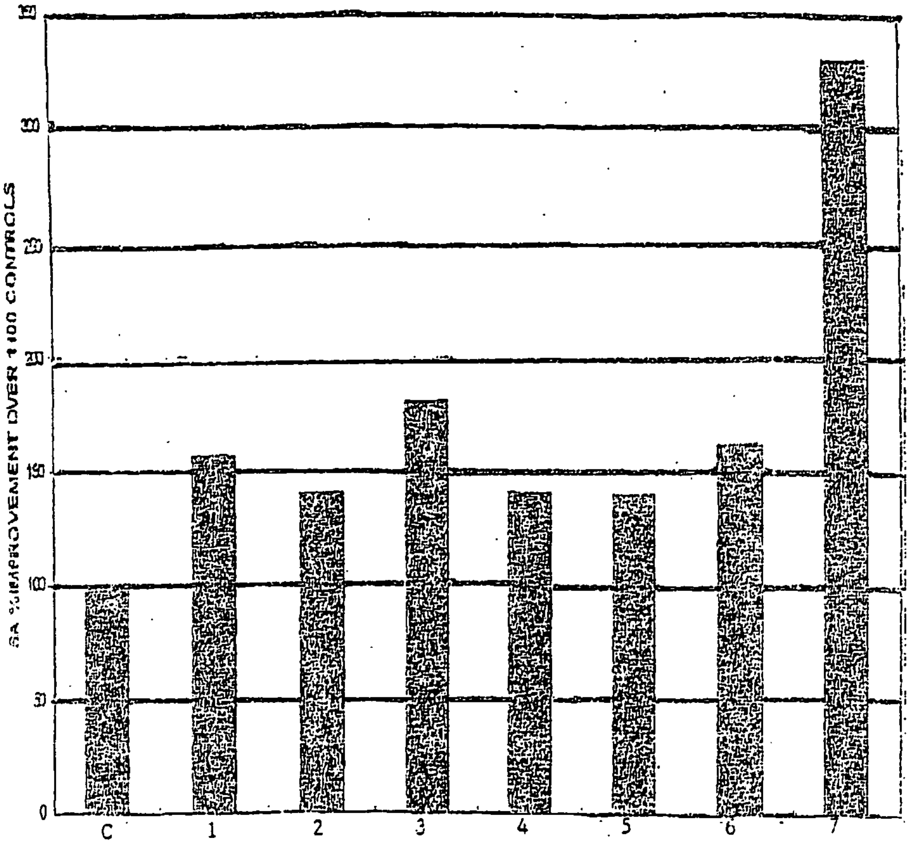


Figure 9

AJK1 dszC: [illegible]
Chimeric dszC: [illegible]
IGTS8 dszC: [illegible]

[illegible]

Figure 10

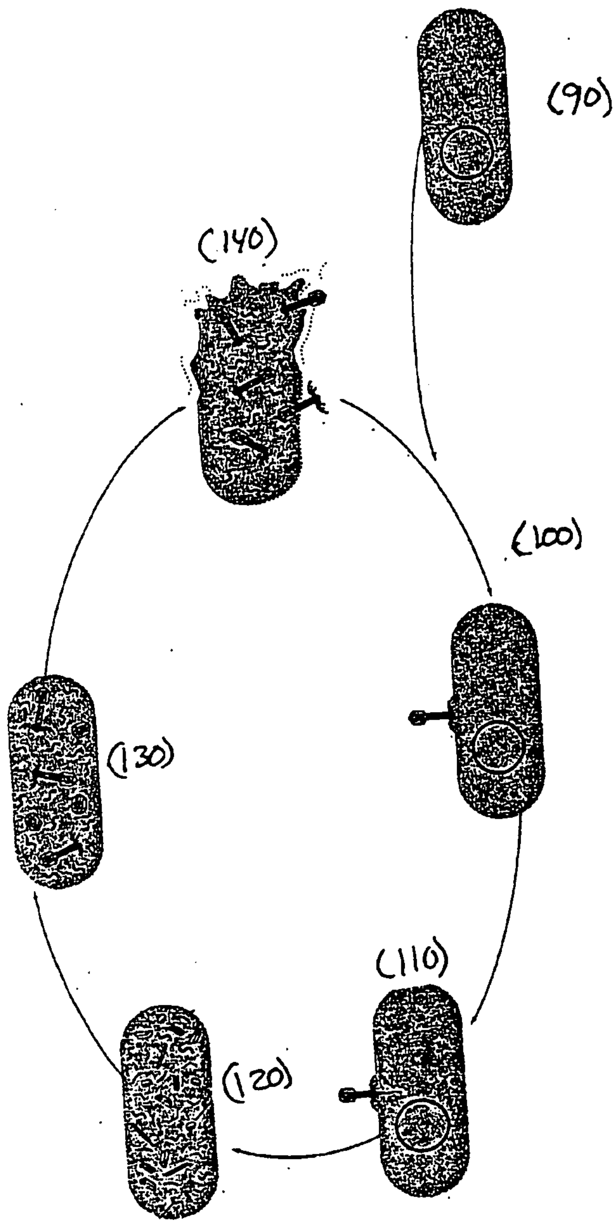


Fig. 11

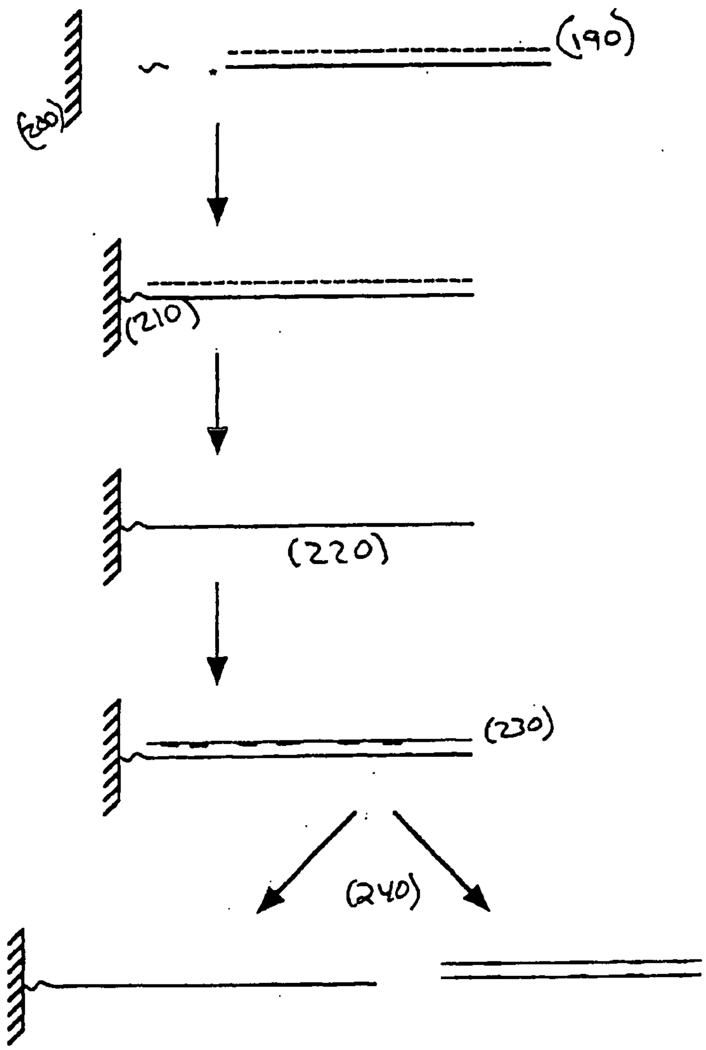


Fig. 12

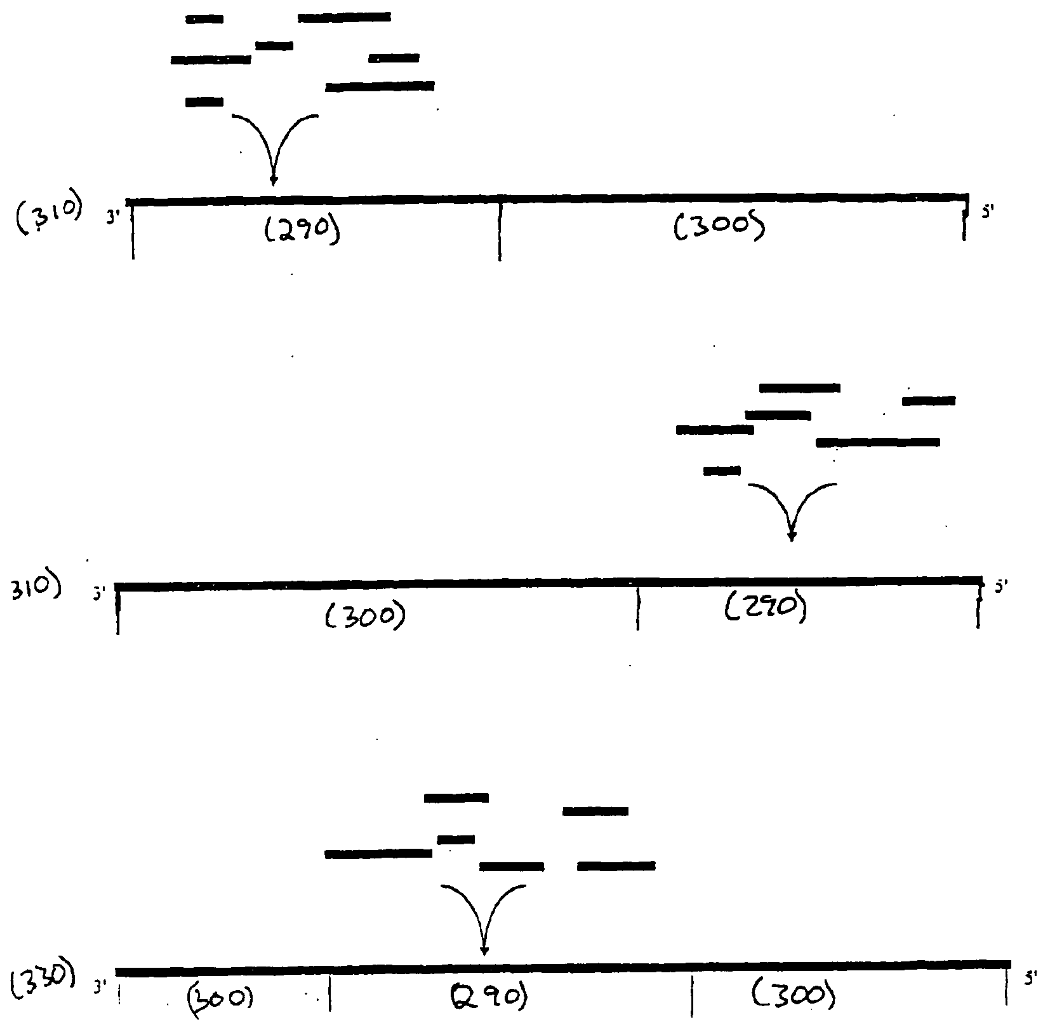


Fig. 13

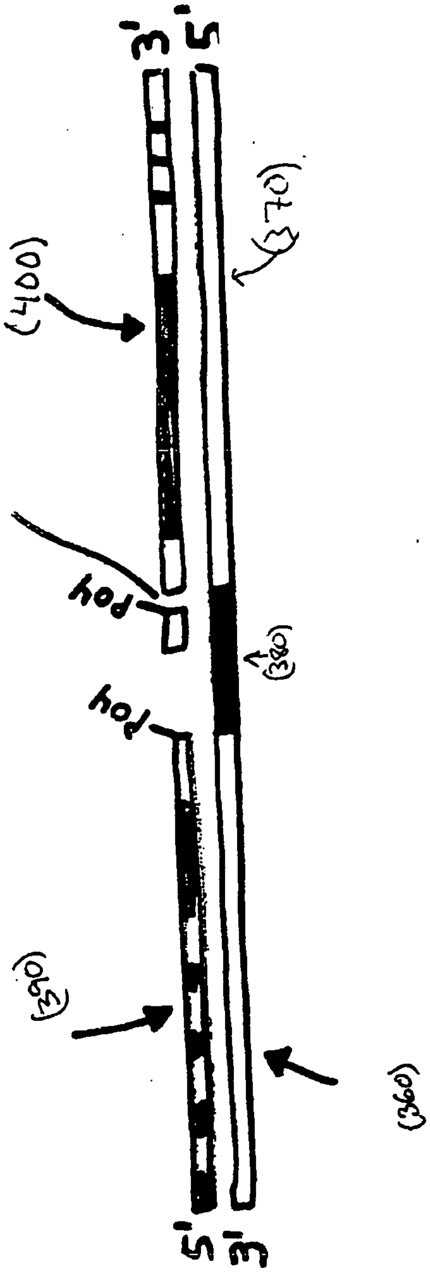


Fig. 14

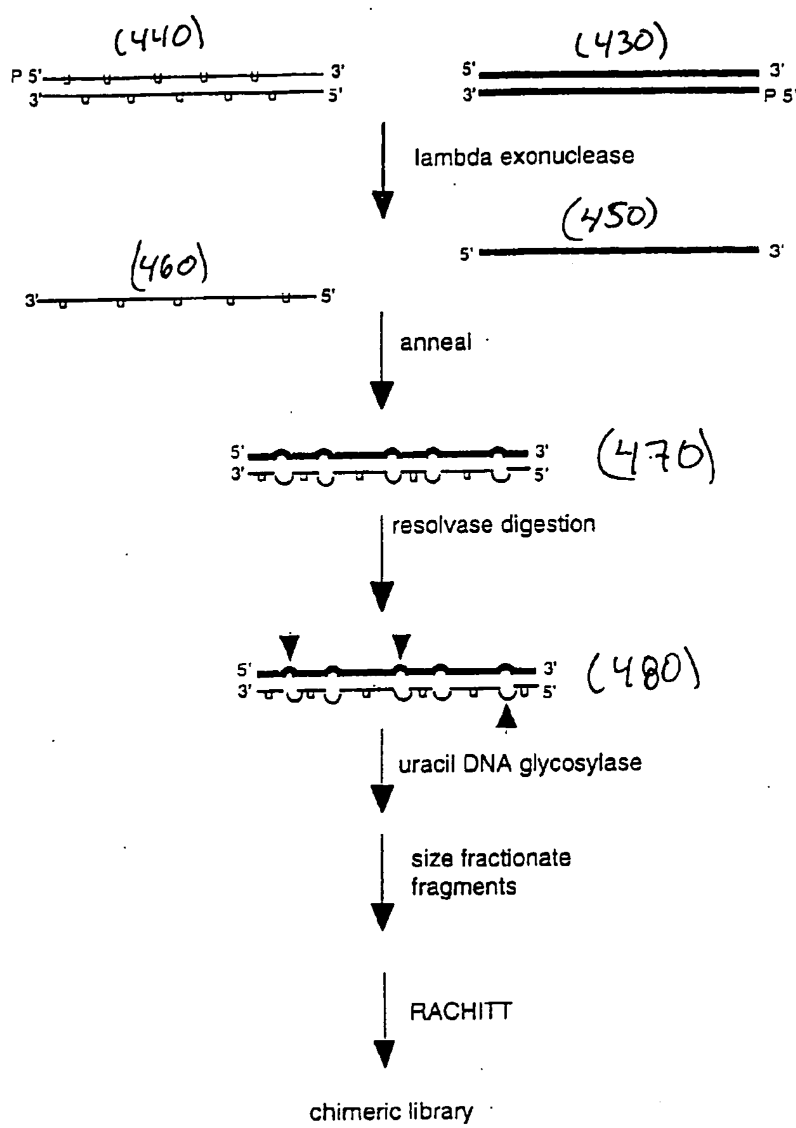

(440)
P 5'
3'
3'
5'
(430)
5'
3'
3'
P 5'
lambda exonuclease
(450)
5'
3'
(460)
3'
5'
anneal
5'
3'
3'
5'
(470)
resolvase digestion
5'
3'
3'
5'
(480)
uracil DNA glycosylase
size fractionate
fragments
RACHITT
chimeric library

Fig. 15

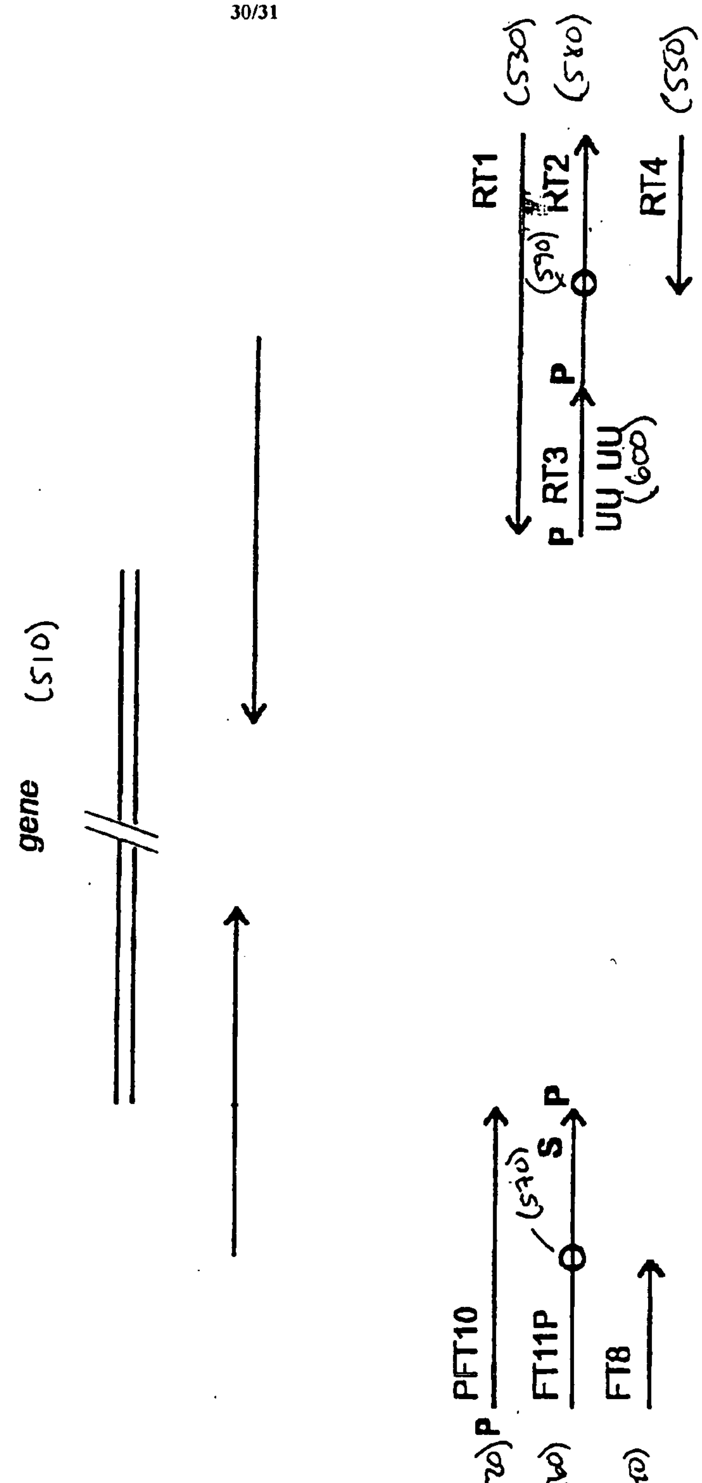

gene
(510)
RT1
(530)
RT2
(580)
(570)
RT4
(550)
P
RT3
UU UU
(600)
PFT10
(520)
FT11P
(560)
(570)
S
FT8
(540)

Fig.16A

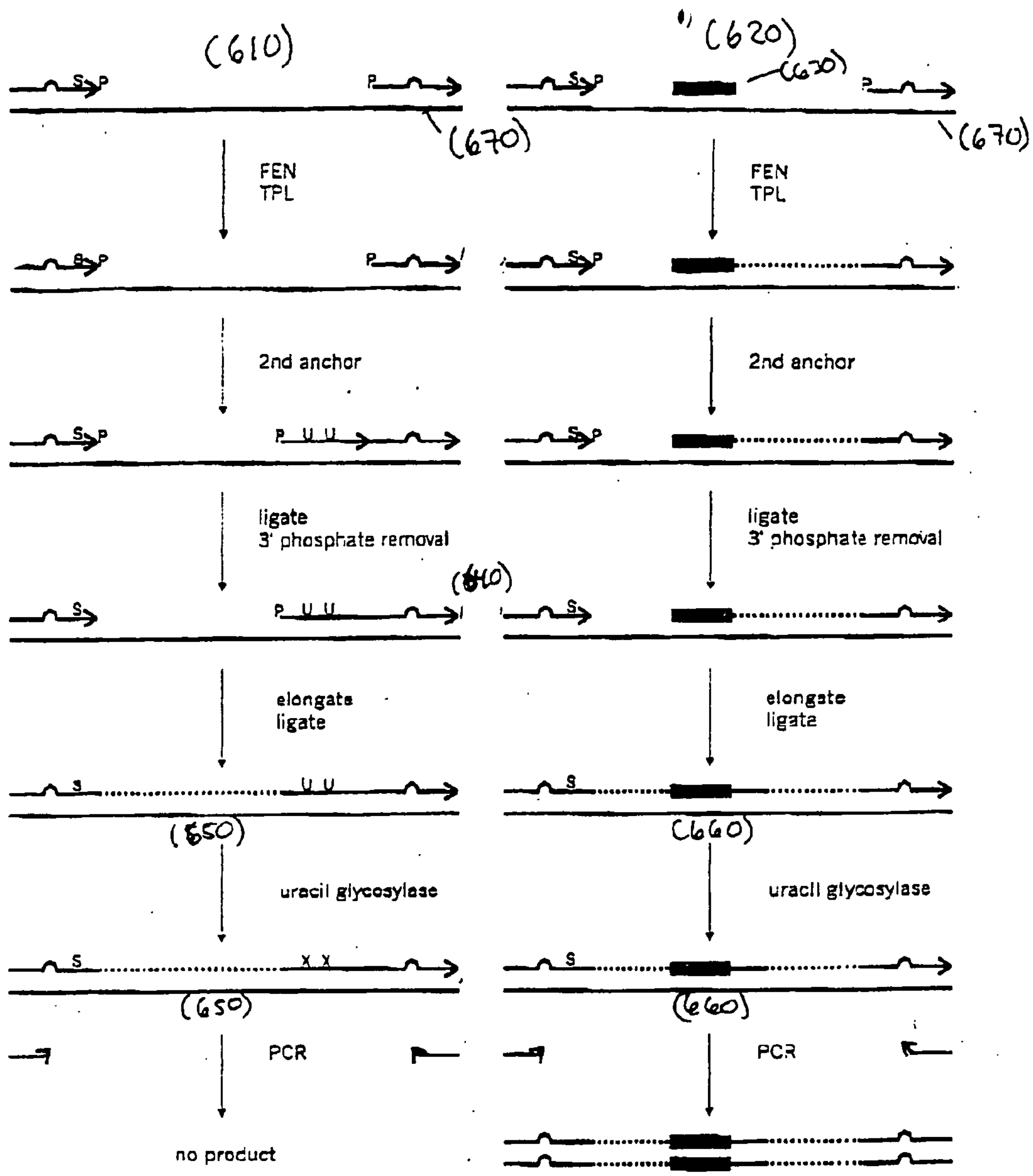

(610)
(620)
(630)
S
P
P
(670)
(670)
FEN
TPL
FEN
TPL
2nd anchor
2nd anchor
P U U
ligate
3' phosphate removal
ligate
3' phosphate removal
(640)
elongate
ligate
elongate
ligate
U U
(650)
(660)
uracil glycosylase
uracil glycosylase
X X
(650)
(660)
PCR
PCR
no product

Fig. 16B